Structure-Function Analysis of Hox-Cofactor Interactions During *Drosophila melanogaster* Embryonic Development.

Katherine M. Lelli

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

© 2012
Katherine M. Lelli
All rights reserved

ABSTRACT

Structure-function analysis of Hox-cofactor interactions during *Drosophila melanogaster* embryonic development.

Katherine M. Lelli

Regulation of gene expression is critical to many aspects of life. From cell survival and proliferation to animal development and species propagation, improper gene regulation can have serious, often fatal, consequences. Therefore, understanding the processes that control gene expression can provide important biological insights. At the center of many of these regulatory processes are trans-acting proteins called transcription factors. Most transcription factors contain DNA-binding domains that recognize specific DNA sequences. These site-specific transcription factors target genes by recognizing binding sites in regulatory sequences called *cis*-regulatory modules (CRMs). However, many transcription factors recognize degenerate DNA-sequences that can be found frequently throughout the genome. Despite this potential for promiscuity, transcription factors control very specific *in vivo* functions. This "specificity paradox" is best understood in the context of one particular family of transcription factors: the Homeobox (Hox) proteins. Conserved in all bilaterians, Hox genes are best known for their roles in embryonic pattering and organogenesis. Characterized by a highly conserved DNA-binding domain called the homeodomain, all Hox proteins recognize similar 'AT' rich sequences. One way Hox proteins achieve functional specificity is through cooperative DNA-binding with the cofactor Extradenticle (Exd) in invertebrates or Pbx in vertebrates. Using *Drosophila melanogaster* as a model

system we conducted a structure-function analysis of three different Hox proteins, Sex combs reduced (Scr), Ultrabithorax(Ubx) and AbdominalA (AbdA) to understand how interactions with a shared cofactor can increase specificity.

To identify amino acid sequence motifs that contribute to Exd-dependent functions, we generated and tested a series of mutant Hox proteins for cooperative DNA-binding ability *in vitro,* and for their ability to regulate target genes *in vivo*. The results of these studies demonstrate that while Scr uses a single conserved motif, more posteriorly expressed Hox proteins Ubx and AbdA use multiple, sometimes unique motifs to regulate Exd-dependent functions. This discrepancy between the quantity and quality of motifs endows AbdA with the ability to outcompete Scr for DNA-binding and regulation of an Exd-dependent target. In addition, by testing the ability for AbdA mutants to carry out a variety of *in vivo* functions, we observed that the different modes of interaction with Exd affect functional specificity. However, in the case of Ubx, we find that despite the contribution of Exd-interaction motifs to cooperative complex formation *in vitro*, none of these motifs are required individually or in combination for *in vivo* functions. Together, these data suggest that one technique Hox proteins use to differentiate themselves when interacting with a shared cofactor is through the utilization of different interaction motifs. Furthermore, having multiple modes of interaction can expand and alter their functional specificity. However, as illustrated by Ubx, the functional interactions between Hox proteins and cofactors can be more complex and may not require cooperative DNA-binding. In conclusion, the characterization of Hox-cofactor interactions helps us better understand how transcription factors select their targets and regulate gene expression in a highly specific manner.

TABLE OF CONTENTS

List of tables and figures.

Acknowledgements

This dissertation is not just the accumulation of six years worth of lab work but is also a testament to all of the people who have helped me along the way.

I'd like to thank my previous undergraduate advisors: Dr Anoop Padival, Dr. Mahin Maines, Dr. Elaine Sia and Dr. David Rempe for nurturing my scientific curiosity and seeing my potential. I'd also like to thank my German professors, especially Dr. Randall Halle, for fostering my creativity and reminding me that there is so much more to the world than science.

I'd like to thank members of the Genetics and Development department for taking a chance on me and giving me this wonderful opportunity. I hope I haven't let you down. Special thanks to all of the administrative staff: Nina Steinberg, Pascale Jean-Louis, Stacy Warren, Celia Marie Morales and Ed Johnson for all their help. I'd like to thank my rotation advisors Dr. Fiona Doetsch and Dr. Iva Greenwald for letting me explore new systems. I'd especially like to thank Iva for continual support both scientifically and personally.

Many thanks to everyone on the 11th floor of Hammer which has been such a great place to work. I have been really spoiled by being able to go just down the hall for reagents and helpful insights. Special thanks to Dr. Gary Struhl who has always taken the time to answer my questions and for the generous use of his confocal microscope. Along with Gary, I'd like to thank the other members of my qualifying and thesis committees: Dr. Lori Sussel, Dr. Oliver Hobert, Dr. Gerard Karsenty and Dr. Licia Selleri. I'd also like to thank Dr. Laura Johnston and members of her lab for interesting discussion and helpful comments during our joint lab meetings. Many thanks to Dr.

Tiffany Cook, Dr. Gary Struhl and Dr. Brian Gebelein for reagents and the Columbia University endocrinology training grant for funding.

As a research scientist your lab is your second home. Very special thanks to all the members of the Mann lab who have been so supportive and understanding. Whether I need virgin females, assistance with a new protocol or someone to grab a beer with I can always count on my lab mates for help. I'd like to especially thank Dr. Barbara Noro for countless hours spent imaging embryos, scoring cuticles and discussing results. Her advice and support these past six years have been invaluable. Most importantly, I'd like to thank Dr. Richard Mann for being such a patient and supportive mentor. It's easy to get overwhelmed by the details but Richard always helped me to see the bigger picture and to find the most interesting aspects of my project.

I am so lucky to have such wonderful and supportive friends and family. Special thanks to all my Delta Gamma sisters, fellow BPG alumni, past roommates and all of my friends at Peridance Center for keeping me sane. Thanks to my fiancé Daniel Concepcion, for intellectual discussions, unconditional love and support. Lastly, I'd like to thank my family, especially my parents, for always believing in me. Without them none of this would have been possible.

To my parents.

Erinnerung.

Willst du immer weiter schweifen?
Sieh, das Gute liegt so nah.
Lerne nur das Glück ergreifen,
Denn das Glück ist immer da.

~Johann Wolfgang von Goethe

CHAPTER 1.

Disentangling the many layers of eukaryotic transcriptional regulation.

Lelli, K. M., Slattery, M., Mann, R.S. Submitted as an invited review to *Annual Reviews of Genetics* April, 2012.

This review was written in collaboration with Matthew Slattery and Richard Mann.

ABSTRACT

Regulating gene expression in eukaryotes is an extremely complex process. In this review we break down several critical steps, emphasizing new data and techniques that have expanded current gene regulatory models. We begin at the level of DNA sequence where *cis*-regulatory modules (CRMs) provide important regulatory information in the form of transcription factor (TF) binding sites. In this respect, CRMs function as instructional platforms for the assembly of gene regulatory complexes. We discuss multiple mechanisms controlling complex assembly including: cooperative DNA-binding, combinatorial codes and CRM architecture. The second section of this review places CRM assembly in the context of nucleosomes and condensed chromatin. We discuss how DNA accessibility and histone modifications contribute to TF function. Lastly, new advances in chromosomal mapping techniques have provided increased our understanding of intra- and interchromomsal interactions. We discuss how these topological maps influence gene regulatory models.

INTRODUCTION

Gene regulation is fundamental for every biological process. Reflective of its importance to cell survival and function, the regulatory mechanisms controlling gene expression are exquisitely sophisticated. Unraveling this complexity has broad implications, from understanding animal development to preventing and treating clinical pathologies. Innovative technologies and creative experimental design are rapidly expanding the current models of gene expression. From new insights into DNA-binding specificity to the contribution of nuclear architecture, this review aims to integrate recent discoveries into a more comprehensive picture of eukaryotic gene regulation.

Gene regulation stripped down to the DNA.

The first gene regulatory model was pioneered by François Jacob and Jacques Monod in the early 1960s. They postulated that gene products are able to feed back and regulate the expression of genes, a concept that laid the foundation for contemporary gene expression models. Currently, the most basic model dictates that regulatory proteins called transcription factors (TFs) act in *trans* to promote or inhibit expression from a locus by binding specific DNA sequences in *cis*- regulatory modules (CRMs) or enhancers (Istrail and Davidson, 2005). TFs are characterized by the sequence and structure of their DNA-binding domains. Throughout evolution gene duplication events have expanded the number of TFs resulting in groups or families of highly related TFs. As a consequence, evolutionarily related TFs often share similar DNA-binding domains and similar in vitro DNA binding specificities. In some cases related TFs display functional redundancy in vivo. However, there are many instances

where individual TFs with highly similar DNA binding properties carry out distinct functions. Given that they have both overlapping and unique functions, related TFs must have the capacity to regulate both common and specific gene targets. While it is easy to understand how TFs with similar binding properties recognize the same binding sites and regulate some of the same target genes, it is less obvious how CRMs restrict binding to specific TFs (Pan et al., 2010). In the first section of this review we focus on recent developments that provide new insights into how specificity is achieved at the level of DNA recognition.

CRMs usually harbor multiple TF binding sites. In some cases, binding sites are only stably occupied by TFs that bind cooperatively. However, the mechanisms by which cooperative DNA-binding increases TF specificity varies according to TF family and CRM (Courey, 2001; Georges et al., 2010; Moretti and Ansari, 2008). Focusing on recent examples with significant structural and functional data, we discuss different ways cooperative binding is achieved. Simply relying on the combinatorial binding and activity of multiple factors is another way CRMs coordinate gene expression, allowing integration of cell type and environmental inputs (Figure 1.1).

Dressing up gene regulation with chromatin.

The biophysical realities of DNA packaged within a nucleus are in stark contrast to the naked DNA researchers typically think about when discussing DNA binding. While the stripped down view is important for understanding the biophysical properties that govern TF-DNA interactions, it ignores all of the complexities of the nuclear environment. DNA in eukaryotic genomes is compacted into chromatin; the basic unit of

chromatin, the nucleosome, consists of 147 base pairs of DNA wrapped around a histone octamer containing two copies of each of the core histones H2A, H2B, H3, and H4 (Li and Reinberg, 2011; Luger et al., 1997). DNA associated with histones is less accessible to TFs and RNA polymerase than naked DNA, making chromatin transcriptionally more repressed compared to naked DNA. Additionally, chromatin structure is not homogenous along the entire genome and can adapt more complex local structures and higher-level three-dimensional arrangements. In the second section of this review we discuss how chromatin structure contributes to gene regulation (Figure 1.1).

Historically, chromatin has been described as existing in two distinct flavors: euchromatin and heterochromatin. Heterochromatin is condensed, transcriptionally inactive, and associated with repressive histone modifications, whereas euchromatin is relatively accessible, and associated with actively transcribed genes and active histone modifications (Felsenfeld and Groudine, 2003; Jenuwein and Allis, 2001). This low-resolution view is still accurate, but recent genomic studies have resulted in a much higher-resolution and more nuanced view of chromatin states. We discuss how nucleosome occupancy as well as histone modifications and variants correlate with gene regulation. Finally, with the help of several technological advances, our understanding of higher-order chromatin structure and the three-dimensional organization of chromosomes in nuclei has increased dramatically over the past decade. We discuss how nuclear architecture and chromosomal conformation have also been implicated in eukaryotic gene regulation (Figure 1.1).

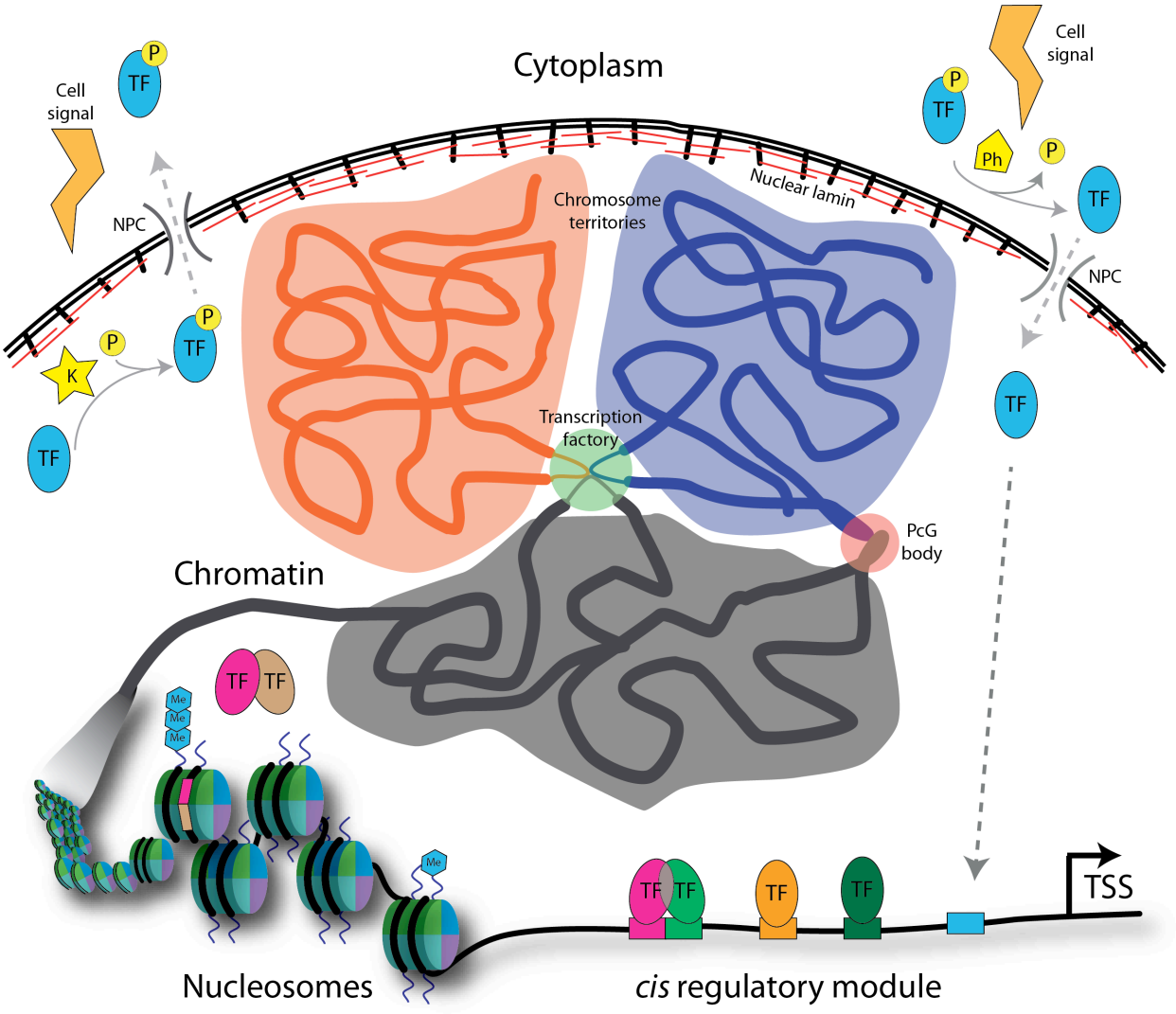


Figure 1.1 Overview of eukaryotic gene regulation.
This cartoon depicts various cellular features that relate to gene expression. TF binding sites are represented as matching colored rectangles. P: Phosphate; K: Kinase; TF: TF; Ph: Phosphatase; NPC: Nuclear Pore Complex; PcG body: Polycomb group body; TSS: Transcription Start Site; Me: methyl group

ASSEMBLING CRM COMPLEXES

In prokaryotes single TFs are able to regulate gene expression. However, this type of gene regulation is insufficient for eukaryotic gene regulation. Instead, eukaryotes rely on combinatorial transcriptional inputs into CRMs to regulate gene expression in space and time (Istrail and Davidson, 2005). The specific recruitment of many individual factors refines expression based on the cellular context, timing of expression and extracellular signals. For example, the same binding sites in the same CRM have been shown to bind different forkhead domain TFs in different tissues, with distinct regulatory outputs (Zhu et al., 2012). On the other hand, multiple homeobox (Hox) TFs, which have highly similar DNA binding specificity as monomers, can target the same gene via distinct CRMs in different tissues (Enriquez et al., 2010). It also appears that TFs can bind non-canonical motifs in certain contexts, though the mechanism by which these motifs are distinguished from canonical motifs remains unclear (Badis et al., 2009; Busser et al., 2012). Regulation can be further refined by post-translational modifications (PTMs) of TFs, which can affect subcellular localization, DNA binding, and protein-protein interactions (Figure 1.1) (Benayoun and Veitia, 2009; Bernard and Harley, 2010; Charlot et al., 2010; Daitoku et al., 2011; Tootle and Rebay, 2005). Some researchers propose a PTM code, where multisite PTM events provide an important regulatory mechanism for different signaling pathways to affect TF function and influence gene expression (Benayoun and Veitia, 2009). In this section we focus on two additional mechanisms, cooperative DNA binding and combinatorial codes that regulate the assembly and activities of CRM complexes.

Cooperativity

One mechanism cells use to increase the DNA-binding specificity of TFs is through cooperative DNA-binding. With an emphasis on structural data paired with in vitro DNA-binding assays, we distinguish three types of cooperative complex formation (Figure 1.2). The first type, which we refer to here as classical cooperativity, relies on direct protein-protein interactions between TFs and their cofactors to increase DNA binding affinity (Figure 1.2). A variation on classical cooperativity, termed 'latent specificity', is when protein-protein interactions not only lead to increased DNA binding affinity, but also to a change in DNA binding specificity (Figure 1.2) (Slattery et al., 2011). We refer to a second form of cooperativity as enhanceosome or modular cooperativity (Figure 1.2). The distinguishing feature here is that, unlike classical cooperativity, which is typically defined for homo- and heterodimers of TFs, enhanceosome cooperativity is observed for large complexes of proteins and, at least in some cases, appears to not depend on protein-protein interactions (Panne, 2008). A third form of cooperativity is termed 'collaborative competition' (Figure 1.2) (Polach and Widom, 1996). In this case, cooperative binding only occurs on a chromatin template because it is the result of multiple TFs being more effective at competing with nucleosome binding compared to individual TFs (Miller and Widom, 2003). As the first two forms of cooperativity are measured on naked DNA we discuss them here, while the third form, which is only apparent in the context of chromatin, will be discussed in a later section.

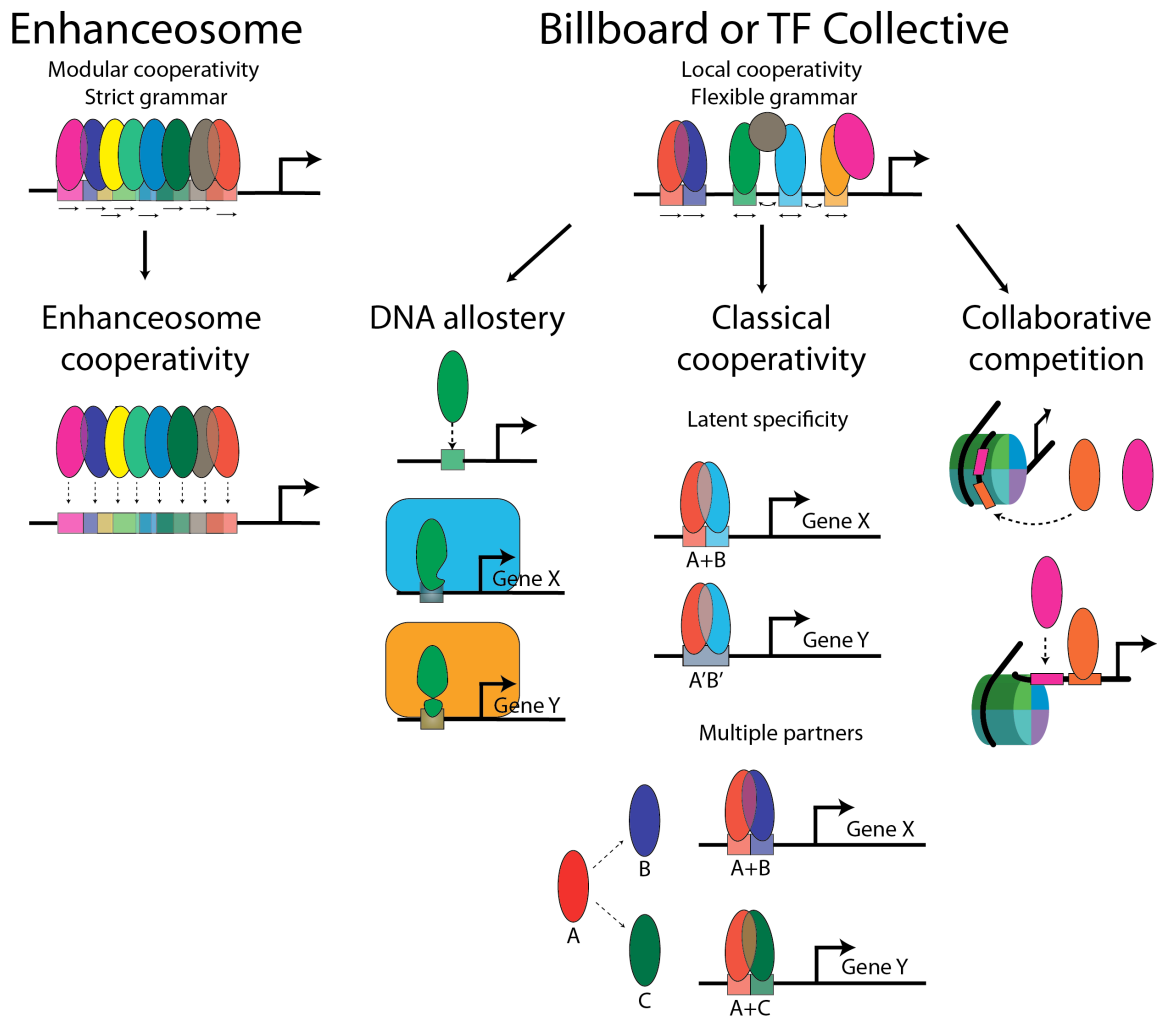


Figure 1.2. CRM assembly and cooperative DNA-binding models.
This cartoon depicts three different models of CRM assembly and related cooperativity mechanisms. (left) The enhanceosome model requires strict modular cooperativity between all TFs (Pan and Nussinov, 2011; Panne, 2008). (right) In contrast, the flexibility of the Billboard and TF Collective models permits different cooperativity mechanisms to control CRM assembly. In the case of DNA allostery, interactions between the DNA sequence and the TF can facilitate conformational changes in the TF that results in the recruitment of different regulatory complexes (depicted as rounded rectangles) in a sequence-specific manner (Meijsing et al., 2009). Classical cooperativity uses protein-protein interactions between TFs to facilitate cooperative binding. These types of cooperative interactions help to increase TF DNA-binding specificity by restricting recruitment to dimeric sites (A+B). In the case of latent specificity, direct protein-protein interactions alter binding specificities so that TFs recognize novel composite sites (A'B') (Slattery et al., 2011). Lastly, collaborative competition between TFs and nucleosomes can lead to cooperative binding when the binding of one TF provides access for another TF to bind a neighboring site (Mirny, 2010; Moyle-Heyrman et al., 2011; Tims et al., 2011).

Protein-protein interactions reveal latent specificities

Interactions between TFs can not only increase DNA binding affinity to their cognate binding sites, but can also result in a modification of their DNA recognition properties (Figure 1.2). This phenomenon has recently been described for the Hox family of TFs, which provide a classic example of the TF specificity paradox. Characterized by a highly conserved DNA-binding domain called the homeodomain, Hox proteins all recognize similar AT-rich sequences in vitro (Berger et al., 2008; Noyes et al., 2008), but confer phenotypically distinct identities in vivo (Hughes and Kaufman, 2002). One way they achieve functional specificity is through cooperative DNA-binding with cofactors. The best-characterized Hox cofactors are, in *Drosophila*, Extradenticle (Exd) and Homothorax (Hth), while in vertebrates they are Pbx and Meis, respectively (Lelli et al., 2011; Moens and Selleri, 2006). Exd/Pbx and Hth/Meis are both members of the three amino acid loop extension (TALE) class of homeodomain-containing proteins (Mann and Chan, 1996; Moens and Selleri, 2006). Functional and structural studies have identified a conserved protein-protein interaction between Exd and Hox proteins that facilitates cooperative DNA-binding (Mann et al., 2009; Merabet et al., 2010). Tryptophan-containing motifs (W-motifs), most commonly YPWM, found N-terminal to the Hox homeodomain, directly interact with the TALE motif in the Exd homeodomain (Mann et al., 2009; Merabet et al., 2010). Mutation of these W-motifs can dramatically affect both cooperative DNA-binding in vitro and in vivo function of some Hox proteins (Lelli et al., 2011; Mann et al., 2009; Merabet et al., 2011; Saadaoui et al., 2011). For example, for two *Drosophila* Hox proteins, Sex combs reduced (Scr) and Deformed (Dfd), mutation of the W-motif was sufficient to abolish cooperative complex formation

on a known specific binding site in vitro, as well as several Exd-dependent functions measured in vivo (Joshi et al., 2010; Lelli et al., 2011). In addition, minimal Hox proteins that contain only the homeodomains and W-motifs retain many wild type functions when assessed in vivo, suggesting that they are sufficient for many in vivo activities (Papadopoulos et al., 2011; Papadopoulos et al., 2010).

In addition to increasing affinity, the binding of Exd with Hox proteins modifies their specificity, a phenomenon referred to as latent specificity (Slattery et al., 2011). Using SELEX-seq, in which traditional SELEX (Tuerk and Gold, 1990) is paired with deep sequencing, cooperative binding with Exd was shown to elicit changes in Hox DNA binding preferences that are distinct from their monomeric DNA binding preferences (Slattery et al., 2011). Importantly, for at least one Hox protein (Ubx), the types of latent heterodimer specificities revealed by SELEX-seq were over-represented in DNA sequences bound by this factor in vivo, arguing that the in vitro-measured specificities are biologically relevant (Slattery et al., 2011). In another recent example, DNA binding measurements from protein-binding microarrays (PBMs) for the *S. cerevisiae* TFs Met4, Met28 and Cbf1 demonstrate that cooperative complex formation increases DNA-binding specificity (Siggers et al., 2011). In this case, Cbf1, together with its non-DNA binding cofactors Met28 and Met4, recognize additional DNA sequences that are adjacent to the traditional Cbf1 binding site (Siggers et al., 2011). Importantly, these additional DNA sequences are essential for gene regulation in vivo. Together, these data suggest that cooperative binding between TFs and their cofactors has the potential to reveal specificities that cannot be utilized in the absence of their cofactors.

Moreover, latent specificities can apparently be induced by both DNA binding (e.g. Hox-Exd) and non-DNA binding (e.g. Cbf1-Met4-Met28) cofactors.

The E twenty-six (Ets) family of TFs may provide an additional example of latent specificity, referred in this case to as acquired specificities (Verger and Duterque-Coquillaud, 2002). Ets proteins are characterized by a highly conserved winged helix-loop-helix DNA binding domain and recognize 5'-GGA(A/T)-3' sequence motifs. Similar to Hox proteins, Ets factors use direct interactions with cofactors to increase DNA-binding specificity (Hollenhorst et al., 2011). Cooperative DNA-binding between PAX5 and ETS1 is critical for activation of *mlb-1* during B-cell development (Fitzsimmons et al., 2001). Structural studies demonstrate that direct interactions between the paired domain of PAX5 and the ETS domain of ETS1 can rearrange protein-DNA contacts to increase DNA-binding (Garvie et al., 2001): a PAX5 induced rotation of a particular tyrosine residue in the recognition helix of ETS1 promotes binding to a low-affinity site (Garvie et al., 2001). A similar mechanism has been proposed for other Ets/cofactor complexes. The ternary complex factor (TCF) subfamily of Ets factors, such as ELK1 and SAP1, all interact with serum response factor (SRF) to activate immediate-early genes (Hollenhorst et al., 2011). In the case of ELK1/SRF complexes, studies of related SAP1/SRF suggest that protein-protein interactions may induce conformational changes in an analogous tyrosine residue in the recognition helix of ELK1 to increase DNA-binding affinity (Mo et al., 2001). Altering the structure of the ETS domain may prove to be a general mechanism for increasing specificity, especially considering it mediates many of the functional interactions between Ets factors and several cofactors (Verger and Duterque-Coquillaud, 2002). An analogous cofactor-induced change in

conformation was also proposed to underlie Hox-Exd latent specificity. In this case, x-ray crystal structures demonstrated that Exd positions a normally unstructured region of the Hox protein Scr so that it can interact with DNA, specifically, a narrow region of the minor groove (Joshi et al., 2007; Mann et al., 2009). The recognition of minor groove structure or, more generally, DNA shape, is widespread among TFs, suggesting that it may be a common mode of DNA recognition (Rohs et al., 2010; Rohs et al., 2009).

The similarities between Ets complexes and Hox complexes extend to another phenomenon called autoinhibition, in which a TF is inhibited to bind DNA due to domains in the TF itself. Mutual relief of autoinhibition has been shown to mediate cooperative complex formation between ETS1 and RUNX1 (Hollenhorst et al., 2011) and for complex formation between Hox and Exd proteins, where the Hox W-motif apparently interferes with monomeric DNA binding (Chan et al., 1996). Therefore, protein-protein interactions can alter DNA-binding specificity and increase affinity by both rearranging protein-DNA contacts and suppressing autoinhibition.

It is noteworthy that some of the DNA interacting residues in the Scr-Exd complex and Ets complexes are outside the traditionally defined DNA binding domain. Although crystal structures are not yet available, the *Drosophila* Hox protein Dfd requires residues that are in an analogous position to Scr's minor-groove interacting residues to bind and regulate some of its specific targets in vivo (Joshi et al., 2010). These observations blur the traditional definition of a DNA binding domain, in that they show that additional motifs can directly contribute to binding when TFs interact with cofactors. Although it is currently not clear which residues in the Cbf1-Met28-Met4 complex are contacting its expanded binding site (Siggers et al., 2011), it is plausible

that, analogous to Hox-Exd, residues not normally considered part of the DNA binding domain are making some of these contacts. Additional NMR or x-ray crystal structures would be very helpful for resolving these questions.

Additional complexity and, perhaps, DNA binding specificity also comes from the fact that some Hox proteins have additional ways to interact with Exd beyond their W-motifs. For example, some Hox proteins, namely Ubx and AbdA, have multiple W- and non-W-motifs that are used to bind DNA cooperatively with Exd/Hth (Chan and Mann, 1993; Lelli et al., 2011; Merabet et al., 2003; Merabet et al., 2011; Merabet et al., 2007; Noro et al., 2011; Saadaoui et al., 2011). Context-dependent interactions between motifs within the same Hox protein have also been proposed to contribute to functional diversity (Merabet et al., 2011; Saadaoui et al., 2011). Unfortunately, no structural information is currently available to show how these additional motifs interact with each other, Exd or Hth. In the case of AbdA, where up to four potential sequence motifs contribute to cooperative complex formation with Exd, some motifs were differentially required depending on the readout examined (Lelli et al., 2011; Merabet et al., 2011; Noro et al., 2011). These data suggest that, depending on the mode of Exd-interaction, different target sites may be recognized and the three-dimensional structure of the bound complex may vary. Having additional interaction modes also contributes to the phenomenon of posterior prevalence, also known as phenotypic suppression, where more posterior Hox proteins dominate in a post-translational manner over more anterior Hox proteins (Duboule, 1991; Duboule and Morata, 1994; Gonzalez-Reyes and Morata, 1990; Gonzalez-Reyes et al., 1990; Mann and Hogness, 1990). A recent study comparing the abilities for Scr and AbdA to bind and regulate a shared target site found

that the quality and quantity of AbdA's Exd interaction modes both increased cooperative DNA-binding affinity in vitro as well as contributed to AbdA's ability to outcompete Scr for target gene regulation in vivo (Noro et al., 2011).

Promiscuous cooperativity with multiple cofactors

As described above, some TFs, such as Hox proteins, increase their specificity by interacting with a small number of cofactors using a variety of mechanisms to reveal latent specificities. Another strategy used by other TFs to expand their regulatory repertoire is by interacting with a large number of protein partners, which allow TFs to gain cell- and tissue-specific control of gene expression depending on the cell-type availability of cofactors (Figure 1.2). For example, Ets factors utilize a variety of partners to bind DNA (Hollenhorst et al., 2011). With more than forty different regulatory partners, the Sox (SRY-related-HMG-box) family of proteins provides another example of how multiple cofactors can contribute to specificity (Bernard and Harley, 2010; Kondoh and Kamachi, 2010). Classified by their HMG (high mobility group) box DNA-binding domain, Sox family proteins control a variety of developmental processes and are key regulators of pluripotency (Bernard and Harley, 2010; Kondoh and Kamachi, 2010). During melanocyte development the SOX10/PAX3 pair activates expression of the TF MITF, which subsequently functions as another SOX10 partner to promote progression of melanocyte differentiation (Bernard and Harley, 2010; Kondoh and Kamachi, 2010). Additionally, recent data suggest that MEF2C can also function as a SOX10 partner to promote maintenance of the melanocyte fate (Agarwal et al., 2011). An analogous

regulatory mechanism is observed during SOX10 regulation of Schwann cell development (Bernard and Harley, 2010; Kondoh and Kamachi, 2010).

Interactions with different partners can also affect DNA-binding specificity, perhaps using mechanisms that are analogous to latent specificity described above. Recently, two studies have demonstrated that single amino acid substitutions in SOX2 and SOX17 can either disrupt or promote, respectively, cooperative binding with OCT4 in vitro. The ability to form cooperative complexes with OCT4 correlated with cell-reprogramming potential (Jauch et al., 2011; Ng et al., 2012). Since most Sox proteins recognize similar binding sites, restricting interactions to only a specific set of cofactors is important for regulating proper CRM binding.

Enhanceosome cooperativity

The CRM responsible for viral-inducible expression of interferon-β (IFN-β) is among the most studied human transcriptional regulatory elements. Eight proteins cooperatively bind this 55 base pair (bp) enhancer in a structure termed the enhanceosome: one ATF2/c-Jun dimer, four interferon response factors (initially IRF-3 which will be replaced with IRF-7 after IFN-β induction) and one NFκB dimer (p50/RELA) (Panne, 2008; Thanos and Maniatis, 1995). Activated by three different pathways, the specific expression of IFN-β is ultimately regulated by the coincidental activation and cooperative binding of all of these factors. Each factor is unable to individually activate IFN-β expression and loss of any single protein abolishes IFN-β activation (Thanos and Maniatis, 1995). Despite the binding sites within this 55 bp element being tightly packed, several crystal structures that capture subsets of the

enhanceosome display a paucity of protein-protein interactions between pairs of dimers (Panne, 2008; Panne et al., 2004, 2007). Additionally, recent molecular dynamics (MD) simulations suggest that the DNA-bound complexes display an unusually high level of flexibility (Pan and Nussinov, 2011). From these observations it seems unlikely that direct interactions between TFs are mediating the observed cooperative DNA-binding. Instead, these studies suggest that sequence-dependent structural changes in the DNA may facilitate binding of TFs to overlapping sites (Pan and Nussinov, 2011; Panne, 2008; Panne et al., 2007). Using a combination of higher and lower affinity sites can regulate the order in which complexes assemble (Pan and Nussinov, 2011); in this way, binding of one factor could facilitate the cooperative binding of another factor to an overlapping low affinity site through complementary structural changes in shared nucleotides. Further, the architectural factor HMGA1a and other secondary factors may also enhance cooperative DNA-binding on the IFN-β enhancer (Panne, 2008). Therefore, the IFN-β enhanceosome represents a type of modular cooperativity with strict requirements for binding site arrangement and overlap (Figure 1.2).

Binding site allostery

Lastly, we wish to emphasize that the binding site, itself, can be an active player in TF function and activity. Studies on the glucocorticoid receptor (GR) have shown that DNA can function as an allosteric regulator of TF activity (Meijsing et al., 2009). Structural studies of the GR demonstrate that a region called the ‘lever arm’ within the DNA-binding domain adopts different conformations according to the DNA sequence bound (Meijsing et al., 2009). These structural changes parallel functional variations in

cognate regulatory complexes for the different binding sites (Meijsing et al., 2009). Therefore, different conformations of the DNA-binding domain induced by the DNA binding site can affect the transcriptional activity of GR in a site-specific manner (Meijsing et al., 2009). As with latent specificity, we speculate that a variety of influences on the three dimensional structure of TF complexes, in part influenced by the DNA binding site, can affect TF functions. Therefore, understanding the sequence and shape of CRMs, and how they impact the structure of bound factors, will be critical to decoding the regulatory logic driving gene expression.

CRM architecture

A common feature of all CRMs is that they provide a scaffold for a combinatorial logic code in which the assembly of multiple factors provides cell type- and environment-dependent gene regulation (Istrail and Davidson, 2005). In the above section, we focused on several types of DNA binding cooperativity used by TFs to bind CRMs. With the exception of enhanceosome cooperativity, where all factors must bind to an inflexible CRM, the other types of cooperativity described above allow for much greater CRM flexibility: a single CRM can potentially integrate a large variety of inputs, some of which may be cooperative. For example, the Hox-targeted CRM from *reaper* integrates not only the Hox protein Deformed (Dfd) but at least eight additional TFs (Stobe et al., 2009). In contrast, a Dfd autoregulatory target requires multiple Dfd-Exd heterodimer inputs plus additional, as yet unidentified, inputs (Joshi et al., 2010). The ability to interchange CRM inputs provides cells with the flexibility to maintain a tight

regulatory control in a wide variety of distinct contexts. Given this requirement, how are CRMs organized, and are there generalizations that can be gleaned from the data?

Three different models have been proposed to describe CRM architecture: the enhanceosome, billboard, and TF collective (Figure 1.2) (Arnosti and Kulkarni, 2005; Junion et al., 2012). As described above, the enhanceosome model posits that cooperative binding of a group of TFs using a strict arrangement or grammar of binding sites in the DNA is necessary for CRM activity (Figure 1.2). Evidence for this model comes primarily from the IFN-β enhanceosome, described above, as well as the Tumor Necrosis Factor-α enhanceosome (Barthel et al., 2003; Panne, 2008). Given the few number of identified enhanceosome-like CRMs, where the arrangement of the binding sites is inflexible, they may be more the exception than the rule. Enhanceosomes may be limited to regulatory events that must be controlled with exquisite precision, as they can only be activated when all factors are present. In addition, once formed, enhanceosomes are unusually stable, allowing them to activate transcription until subsequent mechanisms disassemble them (Thanos and Maniatis, 1995).

A second model for CRM architecture is the billboard model (Figure 1.2) (Arnosti and Kulkarni, 2005). At the opposite end of the spectrum from inflexible enhanceosome CRMs, billboard CRMs are hypothesized to be very flexible, where each binding site is critical, but their relative orientation and spacing do not contribute to CRM function (Arnosti and Kulkarni, 2005; Kulkarni and Arnosti, 2003). According to this view, individual TF inputs might act independently to recruit different components of the basal transcription machinery, adapter complexes, or chromatin modifying complexes to promote transcription (Arnosti and Kulkarni, 2005).

A third view of CRM architecture refers to CRMs as TF collectives (Figure 1.2) (Junion et al., 2012). According to this view, TFs are cooperatively recruited to CRMs, but without a precise motif grammar (Junion et al., 2012). Using whole-embryo chromatin immunoprecipitation (ChIP)-chip and ChIP-seq at different stages during *Drosophila* embryogenesis, binding events for a set of five factors involved in cardiac gene regulation were analyzed (Junion et al., 2012). Interestingly, the authors found that a large number of CRMs included all five factors, and that some combinations were very rare (Junion et al., 2012). For example, when one was missing (Tinman, Tin), the other four factors were rarely found together (Junion et al., 2012). While such observations could represent cooperative binding or cooperative recruitment of these five factors, it could also represent an evolutionary selection for functional CRMs that contain all five inputs, and selection against CRMs that have only four of the five inputs. Another potential explanation for these results is that Tin functions as a so-called pioneer factor that is required to initiate binding to these CRMs in a chromatin environment. Thus, until additional biochemical experiments are carried out, it may be premature to conclude that co-binding of these factors is cooperative, rather than a consequence of selection.

Another feature of the TF collective model is that binding sites need not necessarily be present for every TF present at the CRM (Junion et al., 2012). Accordingly, some factors are indirectly bound at CRMs due to protein-protein interactions, alone (Figure 1.2). While this is certainly plausible, it is also possible that low affinity binding sites or latent specificity mechanisms make it difficult for current computational methods to recognize all of the essential binding sites in CRMs.

Consistent with this idea, low affinity binding sites are critical for the accurate activities of some CRMs (Parker et al., 2011).

Detailed in vivo structure-function analysis of the *sparkling* CRM of *Drosophila Pax2* in eye development provides another informative view of CRM logic (Swanson et al., 2010). In this case, the authors asked how binding sites for the known TFs rearrange relative to each other during the course of evolution (Swanson et al., 2010). Several interesting conclusions come from this work. For one, *sparkling* CRMs from different *Drosophila* species can have a very different arrangement of binding sites; yet still function to drive accurate expression in *D. melanogaster*, similar to conclusions obtained from evolutionary comparisons of the stripe 2 CRM from *evenskipped* (Ludwig et al., 2011; Ludwig et al., 2005). In addition, despite rapid evolutionary turnover of binding sites at the *sparkling* CRM, the authors were able to recognize that the spacing between pairs of some binding sites was conserved (Swanson et al., 2010). This conserved CRM grammar suggests that there may be interactions, either direct or mediated by additional factors, between the bound TFs. The idea that cooperative or interacting subelements could be among a set of otherwise independent inputs was also part of the original billboard model (Arnosti and Kulkarni, 2005), and is consistent with the important role of TF DNA binding specificity discussed in the previous section.

Regardless of whether biologists refer to CRMs as a billboard, a TF collective, or some other nom du jour, the emerging view from numerous studies is that many, perhaps the majority of enhancer elements are not enhanceosome in nature, but instead flexibly integrate multiple TF inputs in a surprisingly large number of arrangements. Consistent with a key role for multiple TF inputs, a recent comparative

study found that in vivo binding sites for the mesodermal TF Twist (Twi) is highly conserved across several *Drosophila* species, and loss of Twi binding in one species is often associated not with loss of a Twi motif, but loss of a cofactor binding site (He et al., 2011). Some of these inputs may be singly bound TFs, others may be cooperatively bound pairs of TFs, and yet others may be interacting indirectly via a third factor (Figure 1.2). Such flexibility in CRM architecture makes the *de novo* identification of CRMs, based solely on DNA sequence, a big challenge for biologists, because the same set of inputs can apparently be encoded in the DNA sequence in many ways.

CHROMATIN STRUCTURE

In the previous section we discussed how TFs recognize and bind target sites in the context of 'naked' DNA. However, to fit within the minuscule confines of the nucleus DNA is wrapped around histones and condensed into chromatin. In addition to limiting TF access, histone-DNA complexes are subject to many PTMs that affect gene expression. Furthermore, chromatin is not uniformly distributed throughout the nucleus, so distant regions of the genome, based on linear DNA sequence, may actually be in close proximity. In this section we describe how modifications of histone-DNA complexes and chromatin architecture contribute to gene regulation. It is not possible to catalog all of the many modifications known to occur on histones; for this, the reader is referred to many recent reviews covering this topic (Bannister and Kouzarides, 2011; Barth and Imhof, 2010; Lee et al., 2010; Rando, 2012; Zhou et al., 2011). Instead, we focus on the small subset of modifications that are correlated with CRMs or CRM activity.

DNA accessibility

DNA accessibility is increasingly recognized as an important variable in gene regulation (Kaplan et al., 2011; Kharchenko et al., 2011; Li et al., 2011; Thomas et al., 2011). Although, it has been difficult to mechanistically test the causal relationship between chromatin structure and gene transcription, many genome-wide studies have demonstrated significant correlations. Through modeling of TF binding and DNase1 sensitivity data, two recent studies revealed that chromatin accessibility has a significant impact on the genome-wide binding patterns of a number of developmental regulatory factors expressed in the *Drosophila* embryo (Kaplan et al., 2011; Li et al., 2011). Within regions of open chromatin, TF binding is primarily determined by sequence specificity (Kaplan et al., 2011). A similar correlation between accessibility and TF binding has been observed in mammalian cells (John et al., 2011) and yeast (Zhou and O'Shea, 2011). Additionally, several genome-wide nucleosome-mapping studies reveal trends suggesting that nucleosome positioning may influence gene expression (Bai and Morozov, 2010). First, promoter regions tend to contain nucleosome-depleted regions (NDRs) (Bai and Morozov, 2010). Second, nucleosomes around transcription start sites are often well organized (Bai and Morozov, 2010). The formation of NDRs is predicted to reveal binding sites and facilitate TF binding at regulatory sequences (Mao et al., 2011). Many mechanisms have been proposed to regulate nucleosome positioning and NDR formation (Bai and Morozov, 2010; Jansen et al., 2012; Khoueiry et al., 2010; Segal and Widom, 2009). In addition to the controversial role of the primary DNA sequence (John et al., 2011; Kaplan et al., 2010a; Kaplan et al., 2010b; Kaplan et al., 2009), ATP-dependent chromatin remodelers such as the Swi/Snf complex and binding

by specific TFs have also been implicated in NDR formation (Bai et al., 2011; Lupien et al., 2008; Tolkunov et al., 2011; You et al., 2011). The formation of NDRs falls under the more general heading of establishing gene- and cell type-specific chromatin architectures, which can include the positioning of nucleosomes at regulatory elements and promoters (Floer et al., 2010; Lomvardas and Thanos, 2002).

Similar to nucleosome depleting factors, so-called pioneer factors have been proposed to prime chromatin environments to initiate subsequent TF binding. Unlike many TFs, pioneer factors such as the FOXA and GATA factors, PU.1, and AP1 have the ability bind their target motifs in a 'closed' chromatin environment (Biddie et al., 2011; Heinz et al., 2010; Magnani et al., 2011; Zaret and Carroll, 2011). Upon binding their target DNA, the pioneer factors can drive local chromatin remodeling and create accessible enhancers for additional TF binding (Magnani et al., 2011; Zaret and Carroll, 2011). This type of synergy has previously been termed nucleosome-mediated cooperativity or collaborative competition (Figure 1.2) (Miller and Widom, 2003; Mirny, 2010; Polach and Widom, 1996). Thermodynamic and in vitro studies have demonstrated that TFs can compete with nucleosomes for DNA binding and by unwrapping or evicting the overlapping nucleosome induce the cooperative binding of another TF (Mirny, 2010; Moyle-Heyrman et al., 2011; Tims et al., 2011). Since protein-protein interactions are not required and the relative arrangement and orientation of binding sites are flexible, collaborative competition may be one explanation for why many CRMs have multiple flexible inputs, without a strict grammar (e.g. the billboard and TF collective models, see above) (Arnosti and Kulkarni, 2005; Junion et al., 2012). Additionally, mathematical simulations suggest that by mediating chromatin

reorganization pioneer factors can increase the steady-state binding of other TFs, a process called assisted loading (Voss et al., 2011).

The *Drosophila* zinc-finger protein Zelda (Zld) may represent another type of pioneer factor. Zld is bound to its target motifs throughout the *Drosophila* genome during the maternal-to-zygotic transition, the point when zygotic transcription commences in the developing embryo (Harrison et al., 2011; Liang et al., 2008; Nien et al., 2011; Satija and Bradley, 2012). Most Zld-targeted regions remain bound by Zld and highly accessible later in embryonic development, and are also targeted by numerous developmental TFs at these later stages (Harrison et al., 2011; Nien et al., 2011). Because Zld is highly associated with the relatively 'open' genome before zygotic transcription begins, it is proposed that rather than reorganizing or opening chromatin, Zld binding prevents nucleosome occupancy and maintains DNA accessibility in certain regions of the genome (Harrison et al., 2011).

Histone modifications

Recent studies have greatly expanded the traditional view that chromatin exists in two states, heterochromatin and euchromatin. The emerging view is that there are many varieties of 'active' and 'inactive' chromatin present in a given cell type (Filion et al., 2010; Roy et al., 2010). Moreover, 'active' chromatin territories are not simply permissive for DNA binding. Two recent genome-wide studies, making use of different techniques (van Steensel and Henikoff, 2000) in different cell types have provided higher resolution views that reveal intricate patterns of histone modifications and DNA accessibility at enhancers and promoters (Filion et al., 2010; Kharchenko et al., 2011;

Roy et al., 2010). Despite experimental differences, and the fact that different chromatin factors were studied, the results of the two studies had a number of similarities. Of note, both studies found that TFs were more likely to bind their DNA motifs in the 'enhancer' chromatin state, a sub-region of active chromatin that is characterized in part by monomethylation of histone H3 on lysine 4 (H3K4me1). Methylation patterns on lysine 4 of histone H3 appear to be closely linked to regulatory enhancers, promoters, and active transcription. Monomethylated H3K4 is associated with enhancers, dimethylated H3K4 (H3K4me2) is associated with both enhancers and promoters or transcription start sites (TSS) of actively transcribed genes, and trimethylated H3K4 (H3K4me3) is only associated with the promoters/TSSs of actively transcribed genes (Barski et al., 2007; Heintzman et al., 2007; Rando and Chang, 2009). The presence of H3K4me1 is strongly enriched on developmental enhancers, but this chromatin modification does not correlate with enhancer activity (Bonn et al., 2012). A similar finding, enhancer marking in the absence of transcriptional activation, has been reported for the H3K4me2 chromatin mark (He et al., 2010).

The list of chromatin modifications with the potential to affect TF DNA recognition extends well beyond methylation of histone H3. Enhancers marked with H3K4me1 are susceptible to both transcriptional repression and activation, and these repressed or activated states usually coincide with additional chromatin modifications. At one end of the spectrum, gene silencing and a repressive regulatory state is associated with trimethylation of H3 on lysine 27 (H3K27me3). H3K27me3 modified nucleosomes are generated by the Polycomb repressive complex 2 (PRC2) and recognized by another Polycomb complex (PRC1) (Simon and Kingston, 2009) . PRC1 silences gene

expression, but the mechanism by which this occurs is unclear (Margueron and Reinberg, 2011; Simon and Kingston, 2009). On the other end of the spectrum, when active, H3K4me1-marked enhancers are often also associated with marks such as H3K79me3 and acetylation of H3K27 (Bonn et al., 2012), but little is known about the role of these modifications in 'active' chromatin.

Despite these links between histone modifications and TF binding, whether TFs can recognize a given modification when targeting the genome remains unclear. However, it is becoming increasingly evident that protein-DNA binding can be influenced by histone PTMs; the TF FOXA1 provides evidence for such a mechanism. FOXA1 can bind both DNA and histones, and is selectively recruited to genomic regions with nucleosomes containing mono- or dimethylated H3K4 (Joseph et al., 2010; Lupien et al., 2008; Sekiya et al., 2009; Sekiya and Zaret, 2007; Serandour et al., 2011). Importantly, FOXA1 chromatin binding is attenuated upon loss of H3K4me1 and H3K4me2. This has been taken to suggest that FOXA1's 'reading' of these chromatin marks influences its genomic DNA targeting, although it is also possible that the loss of these chromatin marks indirectly prevents FOXA1 binding (Lupien et al., 2008; Magnani et al., 2011). Additionally, inducible DNA binding of *Drosophila* Heat Shock Factor (HSF) is also influenced by chromatin state (Guertin and Lis, 2010). HSF binding to heat shock elements (HSEs) is dependent on the presence of a canonical HSE DNA motif. This motif is necessary for binding but not sufficient (i.e. only a fraction HSEs are bound in vivo), so binding specificity is dependent on something other than DNA sequence. Additionally, the contribution of repressive chromatin to HSF binding selectivity is minimal, as most of the unbound HSEs are not associated with the H3K27me3

repressive chromatin mark (Guertin and Lis, 2010). Ultimately, it seems that, aside from the HSE motif, the most significant contributor of heat shock inducible HSF binding is an 'active' chromatin state; prior to heat shock, inducibly-bound HSEs are not depleted of nucleosomes but do contain chromatin modifications that are viewed as active and thought to weaken histone-DNA interactions (H3K4me3 and acetylated histones) (Guertin and Lis, 2010). While it is possible that HSF is interacting with one of the 'active' histone modifications, it is just as likely that these modifications have generated an environment where the weakened histone-DNA interactions are more permissive to HSF binding.

Many questions remain with regard to the functional role of these chromatin modifications and whether something resembling a 'histone code' influences genome-wide TF DNA binding (Gardner et al., 2011; Rando, 2012). Are certain post-translational histone modifications simply indicators of repressive or permissive transcriptional environments? Or do histone modifications play a more active role in helping TFs recognize certain subsets of potential regulatory elements? And, beyond the level of chromatin modifications, the presence of histone variants might also influence nucleosome stability and, therefore, DNA accessibility and TF-DNA interactions (Li and Reinberg, 2011; Talbert and Henikoff, 2010). Although much is still to be figured out, the combinatorial possibilities for both binding and regulatory specificity are daunting.

Chromatin interactions in three dimensions

Until this point we have maintained a two-dimensional, linear view of DNA and chromatin. While this is advantageous for understanding basic principles of protein-

DNA interactions, it omits all of the structural complexities associated with regulating gene expression within the three-dimensional confines of the nucleus. Beginning with the initial observation that chromosomes occupy particular regions within the nucleus (Zorn et al., 1979), many studies suggest that the nucleus is divided into distinct functional domains (Figure 1.1). However, these domains are dynamic making them difficult to characterize (Cavalli, 2007). Nevertheless, new techniques are beginning to yield important insights into nuclear architecture, providing strong support for the idea that interactions between distant chromosomal regions contribute in critical ways to gene regulation.

Nuclear organization

Our understanding of chromosomal structure has increased dramatically over the past decade as a result of advances in chromosomal conformation capture (3C) and 3C-based high throughput technologies (de Wit and de Laat, 2012; Dekker et al., 2002), as well as microscopy-based techniques for studying subnuclear DNA or RNA localization (Eskiw and Fraser, 2011; Lieberman-Aiden et al., 2009; Osborne et al., 2004). A general conclusion from these studies is that the interior of the nucleus is not a uniform compartment. This is not unexpected considering the existence of subnuclear structures such as the nucleolus, but the level of organized structure associated with chromatin is striking. Termed “chromosome territories,” these domains are often further organized depending on gene density and activity (de Wit and de Laat, 2012; Dostie and Bickmore, 2012; Ethier et al., 2011; Sexton et al., 2012; Vaquerizas et al., 2011). Gene-rich regions are found more towards the center of the nucleus, while gene-poor

regions are closer to the nuclear periphery. Although the mechanisms are unclear, rearrangements toward the periphery are proposed to be a consequence of interactions with the nuclear envelope (Zuleger et al., 2011). Additionally, active genes are generally found at the surface of a chromosomal territory, while inactive or repressed genes are buried in the interior (de Wit and de Laat, 2012; Dostie and Bickmore, 2012; Ethier et al., 2011; Vaquerizas et al., 2011). Highly expressed genes have also been observed to reside in foci that have been termed “transcription factories” (Razin et al., 2011). These observations and others have lead to the idea that co-localization may lead to co-regulation (Dai and Dai, 2012). Recent Hi-C data in *Drosophila* further correlate intra- and interchromosmal interactions with transcriptionally active regions while inactive domains remain confined within their respective chromosomal territories (Figure 1.3) (Sexton et al., 2012). Additionally, physical domains of interaction correlate with specific sets of epigenetic marks and are demarcated by insulators (Negre et al., 2010; Sexton et al., 2012).

Chromosomal interactions

Since the original finding that CRMs can be far from the promoters they regulate, scientists have strived to understand how different pieces of the genome communicate with each other. The identification of the β-globin locus control region (LCR) and “gypsy” insulators in *Drosophila* established looping as the predominant paradigm for enhancer-promoter interactions, and there are now many examples of enhancer-promoter communication that occur as a result of looping (Krivega and Dean, 2011). Recent experiments in *Drosophila* directly correlate enhancer-promoter interactions with cell-

type-specific gene expression using a new method called cgChIP (cell-and gene-specific ChIP) (Agelopoulos et al., 2012). Using cell type-specific expression of the bacterial DNA binding protein LacI, enhancers of the gene *Distalless* (*Dll*), tagged with LacI binding sites (*lacO*), were observed to interact with the *Dll* promoter in limb primordia cells, where *Dll* is expressed (Figure 1.3) (Agelopoulos et al., 2012). In contrast, enhancer-promoter communication was not observed in homologous abdominal cells where *Dll* is repressed by Hox proteins, implying a more extended DNA conformation in these cells (Figure 1.3) (Agelopoulos et al., 2012). These results imply that local chromatin structure, in this case, enhancer-promoter communication, varies in a cell-type specific manner. In addition to *cis*-looping models, enhancers can also

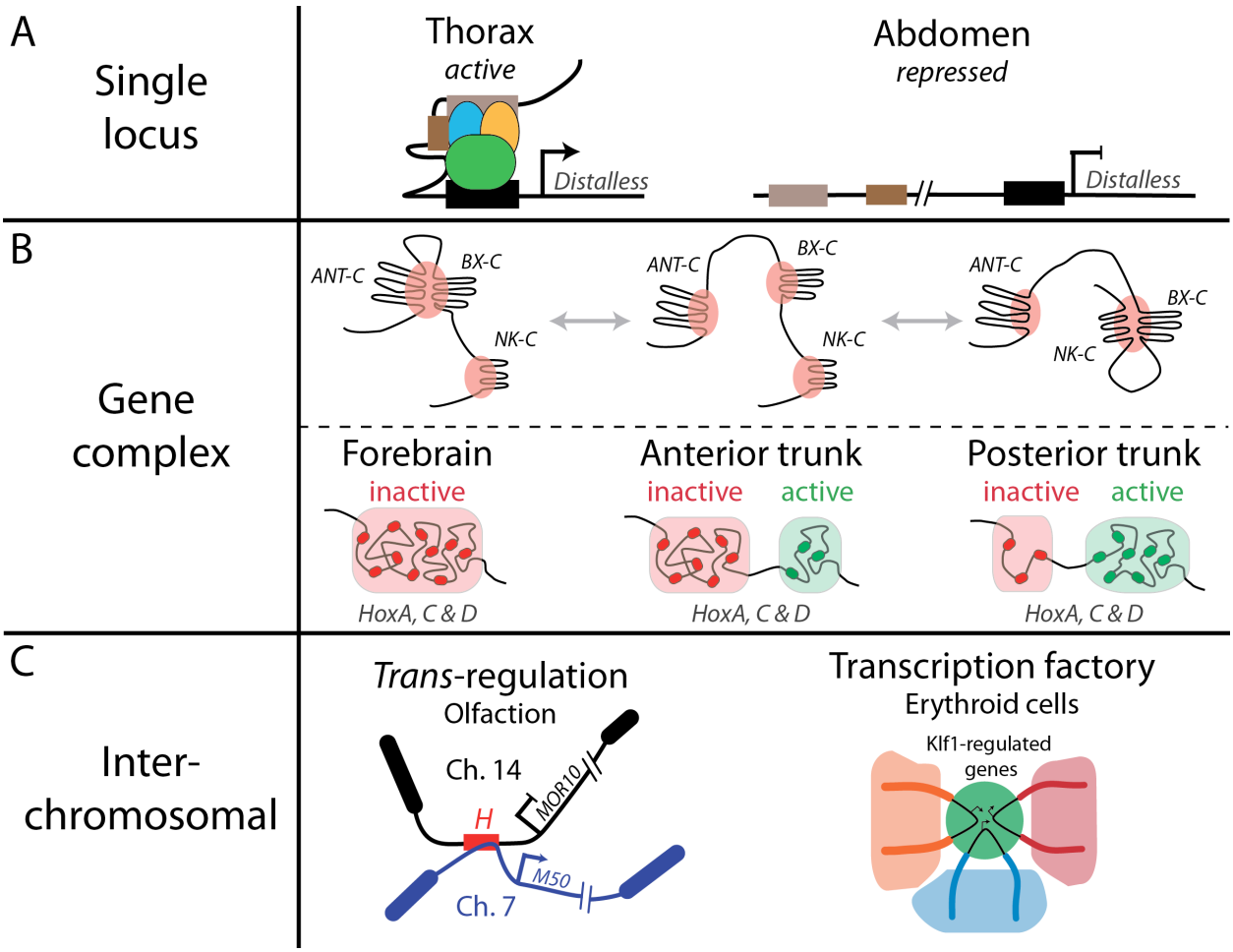


Figure 1.3. Different types of long-range interactions.
(a) Interactions between elements in a single locus. Tissue specific conformations of elements within the *Dll* locus correlate with gene activity. In the thorax where *Dll* is expressed, the locus is more compact with several regions interacting over long distances. However, in the abdomen where *Dll* is repressed, the locus adapts a more extended conformation and no interactions are observed between regulatory elements (Agelopoulos et al., 2012). (b) Interactions between elements in a gene complex. In *Drosophila* three gene complexes: Antennapedia (ANT-C), bithorax (BX-C) and NK homeobox (NK-C) create foci of extensive intra-complex interactions when repressed by polycomb called PcG bodies (Bantignies et al., 2011). Additionally, inter-complex interactions suggest that PcG bodies can encompass more than one complex (Bantignies et al., 2011). However, PcG bodies containing all three complexes have not been observed (Bantignies et al., 2011). During vertebrate development, *Hox* genes cluster according to gene activity (Noordermeer et al., 2011). In the forebrain where none of the *Hox* genes are expressed, 4C analysis indicates that all members of the complex group together (indicated in red). However, in trunk regions *Hox* genes adopt a bimodal distribution where all the active genes (in green) are found in one cluster and the inactive (in red) interact within a different cluster (Noordermeer et al., 2011). Furthermore, this 3D organization correlates with collinear gene activation (Noordermeer et al., 2011). (c) Interactions between chromosomes. During odorant receptor choice in mouse neurons, *trans* interactions can occur between the *H* enhancer and genes on different chromosomes (Lomvardas et al., 2006). In this case we have depicted interactions between the *H* enhancer on chromosome 14 and the *M50* gene on chromosome 7 (Lomvardas et al., 2006). Additionally, interactions between *Klf1*-regulated genes (indicated in green) on different chromosomes have been observed in erythoid cells (Schoenfelder et al., 2010). This colocalization of active genes constitutes a transcription factory.

interact with promoters in *trans*. In *Drosophila* this process is called transvection and has been observed between Hox complexes on homologous chromosomes (Duncan, 2002). In vertebrates, *trans* enhancer-promoter interactions have been observed during odorant receptor (OR) choice in olfactory neurons (Lomvardas et al., 2006). Using 3C, the *H* enhancer on chromosome 14 was observed to interact with OR promoters on chromosomes 14, 7 and 9 (Figure 1.3) (Lomvardas et al., 2006). The possibility for a single enhancer to govern gene expression through interchromosmal interactions

provides one possible mechanism to address the how olfactory neurons choose one out of 1,300 possible ORs (Shykind, 2005).

Further evidence supporting a role for long-range interactions in gene regulation comes from studies on *Hox* clusters. Polycomb-dependent silencing of the Antennapedia and bithorax complexes in *Drosophila* demonstrate intra- and intercomplex interactions (Figure 1.3) (Bantignies et al., 2011). These foci of co-repressed genes by polycomb group (PcG) proteins are called PcG bodies (Figures 1.1 and 1.3) (Bantignies and Cavalli, 2011). Furthermore, removal of participating elements from one complex can weaken gene silencing in the other (Bantignies et al., 2011). These data suggest that PcG body formation is not just a consequence of co-repression but may functionally contribute to gene regulation. Spatial clustering of *Hox* genes is conserved between vertebrates and invertebrates (Ferraiuolo et al., 2010; Montavon et al., 2011; Noordermeer et al., 2011). Recent 3C and 4C analysis of the mammalian *Hoxd* gene cluster in developing limb buds found that functional regulatory regions dispersed within a gene desert upstream of the coding region interacted with the active *Hoxd* promoter (Montavon et al., 2011) (Figure 1.3). These 3D interactions were proposed to form a “regulatory archipelago” that through regulation of *Hoxd* gene expression could function to modulate digit morphology; this concept of partially redundant or shadow enhancers working together to regulate a single gene has been described for multiple *Drosophila* genes (Agelopoulos et al., 2012; Barolo, 2012; Dunipace et al., 2011; Frankel et al., 2010; Hong et al., 2008; Perry et al., 2010). At the *Hoxd* regulatory locus, alterations in chromatin interactions correlated with collinear activation of the *Hox* genes during the development (Noordermeer et al., 2011). In

tissues where the *Hox* clusters are silent, genes were observed to reside in single 3D domain marked by H3K27me3 (Figure 1.3) (Noordermeer et al., 2011). However, once gene expression began a bimodal organization was observed (Figure 1.3) (Noordermeer et al., 2011). Using samples from either anterior or posterior portions of the embryo, genes known to be actively transcribed in those regions occupied one domain that correlated with H3K4me3; while genes known to be silent occupied a separate domain that correlated with H3K27me3 (Noordermeer et al., 2011).

Although these Hox complex studies provide strong correlations, there is some controversy as to the relationship between 3D chromatin structure and gene regulation. In erythroid cells specific intra- and inter-chromosomal interactions between co-regulated genes were dependent on a single TF, Klf1 (Figure 1.3) (Schoenfelder et al., 2010). Additionally, transgenes carrying Klf1-regulated genes relocate to transcription factories when inserted into other genomic locations (Schoenfelder et al., 2010). These results suggest that active, co-regulated genes can preferentially organize into transcriptional interactomes (Schoenfelder et al., 2010). However, other studies using glucocorticoid inducible gene expression in cell lines did not observe significant chromosomal rearrangements upon activation (Hakim et al., 2011). Instead, GR activates genes within pre-existing loci that are enriched for DNase1-hypersensitive sites (Hakim et al., 2011). These studies highlight the functional complexity that can be elucidated using 3C and related strategies. However, they also caution making broad interpretations regarding the role of nuclear architecture in gene regulation as observations can be highly specific to a particular gene or group of genes. Furthermore, interactions, or lack thereof, can be highly cell-type as was recently shown in a

chromosome conformation based study of RNA polymerase II transcription (Li et al., 2012) and by cgChIP in Drosophila (Agelopoulos et al., 2012). New advances in microscopy may help to sort out data from crosslinking-based studies by visualizing interactions in situ (Jones et al., 2011).

Regulatory 'factories' and the hierarchically structured nature of chromatin in general suggest a regulatory environment in the nucleus in which the local concentration of TFs and accessory factors can vary significantly from region to region. Protein concentration is an important determinant of TF-DNA interaction, and has been incorporated into recent models of genome-wide DNA binding (Biggin, 2011; Kaplan et al., 2011; Wunderlich and Mirny, 2009), so the potential for regional variation in TF concentrations throughout the nucleus must be taken into account. Such foci might also lead to the identification of indirect protein-DNA interactions when using crosslinking-based protocol like ChIP; whether this is the case will become evident as more genome-wide 3C and cell-type-specific data become available. Thus, although the study of chromosomal conformation is a relatively nascent field when compared to most of TF biology, these studies have the potential to impact both our interpretation of genome-wide ChIP data and our protein concentration-centric models of genome-wide TF binding.

CONCLUDING REMARKS

Much has been made of the finding that many TFs bind thousands of genomic regions in vivo, perhaps because these numbers easily exceed the number of expected target genes for most sequence-specific TFs (Li et al., 2008; MacArthur et al., 2009;

MacQuarrie et al., 2011). However, the number of binding events is still lower than the predicted number of sites throughout the genome based on DNA sequence alone and close to what is expected when accounting for both DNA sequence and chromatin accessibility (Wunderlich and Mirny, 2009). TFs are most typically viewed as components of 'discrete' regulatory networks, with separate target and non-target genes for each TF (Davidson, 2010; Davidson and Levine, 2008; Negre et al., 2011). However it has been proposed, based on genome-wide binding data, that TF regulatory networks should instead be viewed as 'continuous' networks (Biggin, 2011). The continuous network model of TF function posits that due to the high nuclear concentration of most expressed TFs, specificity mediated by protein-protein interactions is unnecessary, and essentially all genes are targeted by all TFs. According to this view, biological function is determined by quantitative differences in TF binding at accessible DNA rather than binary on/off TF binding (Biggin, 2011). Aspects of this model are supported by data from a survey of *Drosophila* TFs expressed early in *Drosophila* embryogenesis (Kaplan et al., 2011; Li et al., 2011). However, the overexpression of many TFs, for example by the Gal4-UAS method in *Drosophila*, rarely leads to aberrant phenotypes; instead, higher than normal levels of TFs in cells typically results in wild type readouts arguing that other factors besides concentration must be limiting for TF function. Moreover, there are many examples in which small differences in TF DNA binding domains are important for their specific in vivo functions, arguing that DNA binding specificity is critical. In addition, the continuous network model assumes that there are not cell-type specific differences in binding, and that the signal generated by ChIP is indicative of direct DNA binding. But this may not always be the case, as there is ample evidence for

ChIP signals resulting from indirect DNA binding via protein-protein interactions or interactions between regulatory elements (Agelopoulos et al., 2012; Gordan et al., 2009; Heldring et al., 2011; Moorman et al., 2006). Based on the current data, we suggest that chromatin accessibility is critical for limiting which TF binding sites and which CRMs are available in specific cell types, but within accessible regions DNA binding specificities of TFs and TF complexes are essential for determining which binding sites are productively bound within these accessible regions.

The era of TF genomics has clearly changed our view of how TFs target the genome, but many of the methods routinely used are low-resolution or are blind to cell-type specific differences. New, higher-resolution technologies will undoubtedly lead to refinement and restructuring of these models. For example, a new variation on ChIP, termed ChIP-exo, can generate genome-wide TF-DNA binding profiles down to single base resolution (Rhee and Pugh, 2011). Not only does this high-resolution method provide a more precise view of a TF's DNA binding motifs, it also eliminates a significant number of false positive binding events generated by traditional ChIP-chip or ChIP-seq (Rhee and Pugh, 2011). This method has the potential to refine the models of genome-wide TF binding that have been generated over the past five years.

Beyond advances in ChIP, increasingly sophisticated methods for monitoring nuclear organization will begin to generate a comprehensive picture of nuclear and chromosomal structure in vivo. This will require a combination of both 3C-based techniques and super-resolution microscopy techniques, such as stochastic optical reconstruction microscopy (Rust et al., 2006). Combining these approaches with high-resolution ChIP data will be essential for understanding TF-DNA interactions in the

context of chromatin looping and transcription factories. Ultimately, integrating these multiple layers of data, in combination with the studies described in this review, will allow TF biologists to generate and test models regarding direct versus indirect, specific versus nonspecific, and functional versus nonfunctional binding.

CHAPTER 2.

Competition for cofactor-dependent DNA binding underlies Hox phenotypic suppression.

Noro B*, Lelli KM*, Sun L, and Mann RS. *Genes and Development.* November 15, 2011 (25) 2327-2332.
**equally contributing first authors*

Barbara Noro began this project while she was a graduate student as part of her dissertation research. Initial experiments conducted by her and Liping Sun demonstrated that the *in vivo* competition system would be useful to study phenotypic suppression. These early studies produced Figure 2.1 of which, with the exception of the KD measurements, I have also repeated and confirmed. She also constructed and characterized several of the AbdA mutants including: $2W^{Ala}$, ΔC^{263}, ΔC^{220} and ΔC^{197}. Based on her results I constructed and characterized two additional mutants $\Delta^{200\text{-}220}$ and $2W^{Ala}\ \Delta^{200\text{-}220}$. The data provided in all of the figures is a cumulative representation of our combined data with the following exceptions: Liping Sun conducted the EMSAs for the K_d calculations in Figure 2.1, Barbara produced the alignments in Figure 2.2 and I conducted the clonal misexpression experiments in Figure 2.5. Lastly, Richard Mann wrote the manuscript and provided helpful comments and suggestions during experimental design and data analysis.

ABSTRACT

Hox transcription factors exhibit an evolutionarily conserved functional hierarchy, termed phenotypic suppression, in which the activity of posterior Hox proteins dominates over more anterior Hox proteins. Using directly regulated Hox-targeted reporter genes in *Drosophila*, we show that posterior Hox proteins suppress the activities of anterior ones by competing for cofactor-dependent DNA binding. Further, we map a motif in the posterior Hox protein Abdominal-A (AbdA) that is required for phenotypic suppression and facilitates cooperative DNA binding with the Hox cofactor Extradenticle (Exd). Together, these results suggest that Hox-specific motifs endow posterior Hox proteins with the ability to dominate over more anterior ones via a cofactor-dependent DNA binding mechanism.

INTRODUCTION

The Hox or homeotic genes encode a conserved set of homeodomain-containing transcription factors that control morphological identities along the anterior-posterior (AP) axes in both vertebrates and invertebrates. Although this Hox gene function is conserved across phyla, more recently, these genes have been shown to play additional roles in animal development, from motor neuron identity determination to stem cell maintenance (Dasen and Jessell, 2009; Ernst et al., 2004).

One of the hallmarks of the Hox gene family is that individual members are expressed at specific AP positions in the developing embryo. Hox expression patterns are collinear with their positions along the chromosome: Hox factors that are expressed anteriorly are located 3’ to more posteriorly expressed Hox genes. Hox genes are

typically clustered in metazoans; in mice and humans, 39 Hox genes reside in one of four Hox complexes, while in *Drosophila melanogaster* eight genes reside in two complexes. The more anteriorly expressed Antennapedia complex includes *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex-combs reduced* (*Scr*) and *Antennapedia* (*Antp*), which establish the identities of parasegments (PS) 1 through 5; the Bithorax complex codes for the most posteriorly expressed abdominal Hox genes *Ultrabithorax* (*Ubx*), *Abdominal-A* (*AbdA*) and *Abdominal-B* (*AbdB*), required for PS 6 through PS 14 identities (Hughes and Kaufman, 2002).

The collinear distribution of Hox genes also correlates with the ability of posterior family members to functionally dominate over more anterior ones via post-transcriptional cross-regulatory interactions. This phenomenon, known as phenotypic suppression or posterior dominance, was first recognized by analyzing mutants of the *Polycomb* group of genes (Struhl, 1983) and from experiments in which Hox proteins were mis-expressed during *Drosophila* embryogenesis. It was noticed that posterior segments were generally not transformed, even when more anterior Hox proteins were ubiquitously expressed at high levels. For example, ubiquitous expression of Ubx, which normally establishes the PS6 identity, transformed all thoracic and head segments towards this identity (Gonzalez-Reyes and Morata, 1990, 1991; Mann and Hogness, 1990). However, Ubx was unable to transform abdominal segments toward PS6. Similarly, ubiquitous Antp could efficiently transform segments anterior to PS4, where it is normally active, but was unable to transform more posterior regions (Gibson et al., 1990; Schneuwly et al., 1987). Analogous observations have been made in vertebrates, suggesting that this phenomenon is evolutionarily conserved (Bachiller et al., 1994;

Duboule, 1991; Duboule and Morata, 1994). However, the molecular mechanisms responsible for phenotypic suppression are poorly understood.

Several mechanisms have been invoked to account for phenotypic suppression. One way in which posterior Hox factors dominate over anterior Hox proteins is that the former are, in general, transcriptional repressors of the latter; for example, *Ubx* is a repressor of *Antp* (Carroll et al., 1986; Hafen et al., 1984; Harding et al., 1985; Struhl and White, 1985). Post-transcriptional regulation of Hox expression by microRNAs (miRs) has also been postulated to partly underlie phenotypic suppression (Chopra and Mishra, 2006; Singh and Mishra, 2008; Yekta et al., 2008). According to this model, miRs that target anterior Hox genes are expressed in more posterior segments, and thus would be available to suppress anterior Hox gene functions. Consistent with this idea, miRs that are predicted to target more anterior Hox genes are often located 5' to that Hox gene and, due to collinearity, would be expected to be expressed in more posterior domains. Consistently, several *Drosophila* miRs have been identified that have the predicted ability to suppress more anterior Hox gene functions when ectopically expressed (Ronshaugen et al., 2005).

However, these mechanisms cannot fully explain phenotypic suppression, which functions, at least in part, at the post-translational level. For example, when either Ubx or Antp is expressed ubiquitously, they fail to transform abdominal segments despite high levels of expression in all embryonic cells (Gonzalez-Reyes et al., 1990); the abdominal Hox factors AbdA and AbdB post-translationally block Ubx and Antp functions in these segments. Moreover, when Ubx and Antp are forced to be co-expressed in all embryonic cells, anterior segments are transformed towards PS6, not PS4,

demonstrating that the activity of Antp is suppressed in a post-translational manner by Ubx (Gonzalez-Reyes et al., 1990). In general, the transgenes used to express these Hox proteins do not contain the native Hox gene 3' UTRs that are typically targeted by miRs. Moreover, mutants that delete Bithorax complex miRs do not display any of the homeotic transformations that would be expected if posterior dominance was compromised (Bender, 2008). Together, these observations suggest that other post-translational mechanisms are at play.

Here we test two alternative models that can account for the post-translational nature of phenotypic suppression. Hox DNA binding affinity and specificity are often enhanced by cooperatively binding with cofactors, such as the homeodomain proteins Extradenticle (Exd) in *Drosophila* and Pbx in vertebrates (Mann et al., 2009; Moens and Selleri, 2006). This raises the possibility that competition for cofactors – either on or off DNA – could underlie phenotypic suppression. To test these models, we developed an *in vivo* assay for phenotypic suppression using two Hox factors that have distinct activities: Scr, an anterior Hox protein, and AbdA, a posterior Hox protein. Using two well-defined reporter genes that are dependent on Hox-Exd binding sites for their regulation, we show that AbdA can only suppress Scr's activity when it can compete for DNA binding in a cofactor-dependent manner. In addition, we identified an evolutionarily conserved sequence motif in AbdA that enables cooperative complex formation with Exd and is required for AbdA's ability to dominate over anterior Hox proteins. Together, these results suggest that phenotypic suppression occurs when anterior and posterior Hox proteins compete for the same binding sites in shared target genes. In such cases,

the unique molecular architecture of posterior Hox factors results in a higher affinity for these targets, thus imposing their functions even in the presence of anterior Hox factors.

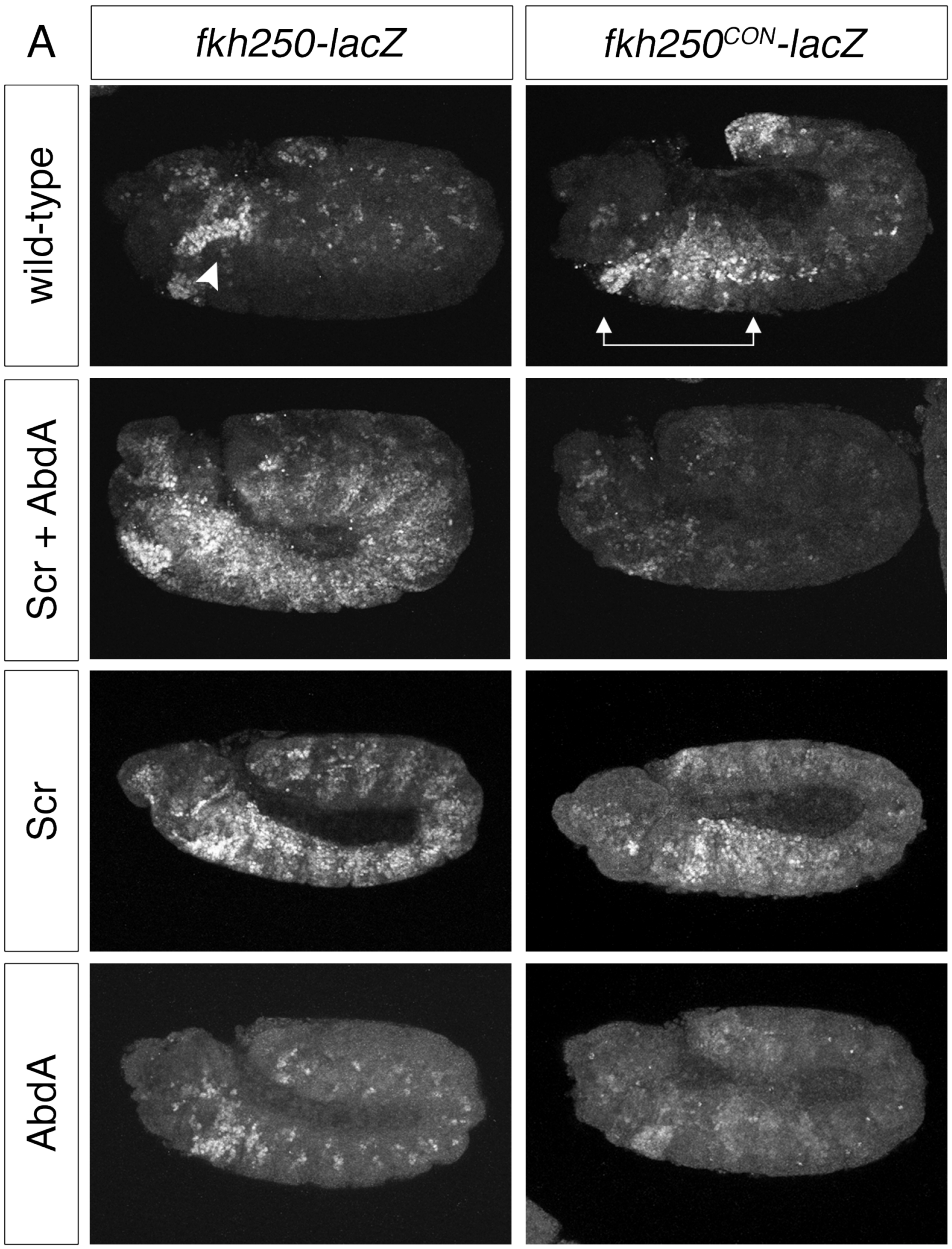
RESULTS AND DISCUSSION

fkh250 is a 37 base pair element from the *forkhead* (*fkh*) gene, which is directly regulated by Scr; it contains a single Hox-Exd binding site that, compared to other Hox-Exd heterodimers, is preferentially bound by Scr-Exd *in vitro* (Ryoo and Mann, 1999). When *lacZ* is placed under the control of *fkh250, fkh250-lacZ* is specifically expressed in PS2 in an *exd*- and *Scr*-dependent manner (Ryoo and Mann, 1999). Indeed, mis-expression of Scr throughout the *Drosophila* embryo can ectopically activate *fkh250-lacZ* (Figure 2.1A) (Joshi et al., 2007; Joshi et al., 2010; Ryoo and Mann, 1999). Notably, ectopic activation of *fkh250-lacZ* occurs even in the abdomen, in the presence of endogenous, more posterior Hox.

In contrast to *fkh250*, *fkh250^{CON}* (for 'consensus') is an artificial variant of *fkh250* with two base pair substitutions that enable *fkh250^{CON}-lacZ* to be directly regulated by four Hox genes in an *exd*-dependent manner: *Scr, Antp,* and *Ubx* activate this reporter in PS2 to PS6, while *AbdA* represses it in abdominal segments (Figure 2.1A and Ryoo and Mann, 1999). Consistent with its relaxed specificity *in vivo, fkh250^{CON}* binds well to Scr-Exd, Antp-Exd, Ubx-Exd and AbdA-Exd heterodimers *in vitro* (Ryoo and Mann, 1999). The promiscuous binding and regulation by multiple Hox proteins classifies *fkh250^{CON}* as a shared Hox target gene, while the Scr-specific regulation of and binding to *fkh250* suggests that it is a specific Hox target gene (Mann et al., 2009).

Because of their distinct specificities, *fkh250-lacZ* and *fkh250^{CON}-lacZ* provide an ideal system to examine the molecular mechanism of phenotypic suppression. In accordance with the premise of posterior dominance, co-expression of Scr and AbdA throughout the fly embryo leads to repression of *fkh250^{CON}-lacZ* by AbdA (Figure 2.1A). In contrast, *fkh250-lacZ* is robustly activated by Scr even in the presence of AbdA (Figure 2.1A). Note that both *fkh250* and *fkh250^{CON}-lacZ* require direct binding by the Hox cofactor Exd (Ryoo and Mann, 1999). The primary distinction between these two readouts is that AbdA-Exd binds well to *fkh250^{CON}* but not to *fkh250*. Accordingly, we conclude that AbdA cannot suppress the activities of Scr if it cannot bind to the target element. Further, in this system, posterior dominance cannot be mediated by miR activity or competition for factors, such as Exd, off DNA. Rather, these data support a model in which competition for cofactor-dependent DNA binding underlies phenotypic suppression for shared Hox target genes.

Figure 2.1. AbdA dominance over Scr relies on an Exd-dependent DNA binding mechanism.
(A) Embryos carrying *fkh250-lacZ* (left panels) or *fkh250^{CON}-lacZ* (right panels) stained for β-galactosidase (β-gal). In wild-type embryos, endogenous Scr activates *fkh250-lacZ* in PS2 (arrowhead), while *fkh250^{CON}-lacZ* is activated by Scr, Antp and Ubx from PS2 to PS6 (arrows) and repressed by AbdA in the abdominal segments. Ectopic expression of Scr throughout the embryo, alone or in combination with AbdA, results in widespread activation of *fkh250-lacZ*. In contrast, *fkh250^{CON}-lacZ* is repressed when Scr and AbdA are both ectopically expressed. (B) Representative *in vitro* saturation binding experiments and dissociation constant (K_d in nM) fits are shown for Scr-Exd (left) and AbdA-Exd (right) binding to *fkh250^{CON}*. AbdA-Exd dimers bound more tightly to *fkh250^{CON}* than Scr-Exd did, supporting a model in which AbdA dominance depends on cofactor-dependent DNA binding. All assays were performed in the presence of Exd-Hth^{HM}. The reported K_ds represent the averages and standard error of the means of repeated measurements (n=5 for AbdA and n=6 for Scr; see Materials and Methods).


A
fkh250-lacZ
fkh250CON-lacZ
wild-type
Scr + AbdA
Scr
AbdA
B
fraction bound
Kd = 25.4 ±1.5
n = 6
Scr concentration (nM)
fraction bound
Kd = 10.6 ±1.9
n = 5
AbdA concentration (nM)

If AbdA is out competing Scr for binding to *fkh250^{CON}*, AbdA would be expected to have a higher affinity for this sequence compared to Scr. To test this prediction, we measured the affinities of AbdA-Exd and Scr-Exd heterodimers for *fkh250^{CON} in vitro*. AbdA-Exd heterodimers bound over two-fold more tightly to *fkh250^{CON}* compared to Scr-Exd (with dissociation constants, K_d, of 10.6 ± 1.9 nM and 25.4 ± 1.5 nM, respectively, Figure 2.1B). Thus, at the same concentration, AbdA-Exd is more likely than Scr-Exd to be bound to *fkh250^{CON}*, consistent with the idea that competition depends on cofactor-dependent DNA binding.

Binding to *fkh250^{CON}* is Exd-dependent for both AbdA and Scr (Ryoo and Mann, 1999), implying that AbdA has a domain, or domains, that allow higher binding affinity with Exd to this target site. In general, Hox interactions with Exd are mediated by the highly conserved, four amino acid motif, YPWM, which directly binds to a hydrophobic pocket established by the three amino acid loop extension (TALE) in the Exd homeodomain (Mann et al., 2009). For some Hox proteins, the YPWM-TALE interaction is necessary and sufficient for cooperative DNA binding with Exd and target gene regulation *in vivo* (Joshi et al., 2010). In addition to the YPWM motif, AbdA, but not Scr, has a second well-conserved tryptophan-containing motif, TDWM, which could play a role in mediating AbdA-Exd interactions (Figure 2.2). However, when a mutant form of AbdA in which both the YPWM and TDWM motifs are mutated to alanines (2W^{Ala}) was co-expressed with Scr in our phenotypic suppression assay, *fkh250^{CON}-lacZ* was repressed to the same extent as by wild-type AbdA (Figure 2.3C, D). Thus, although the YPWM and TDWM may contribute to interactions with Exd, these motifs are not necessary for AbdA to dominate over Scr.

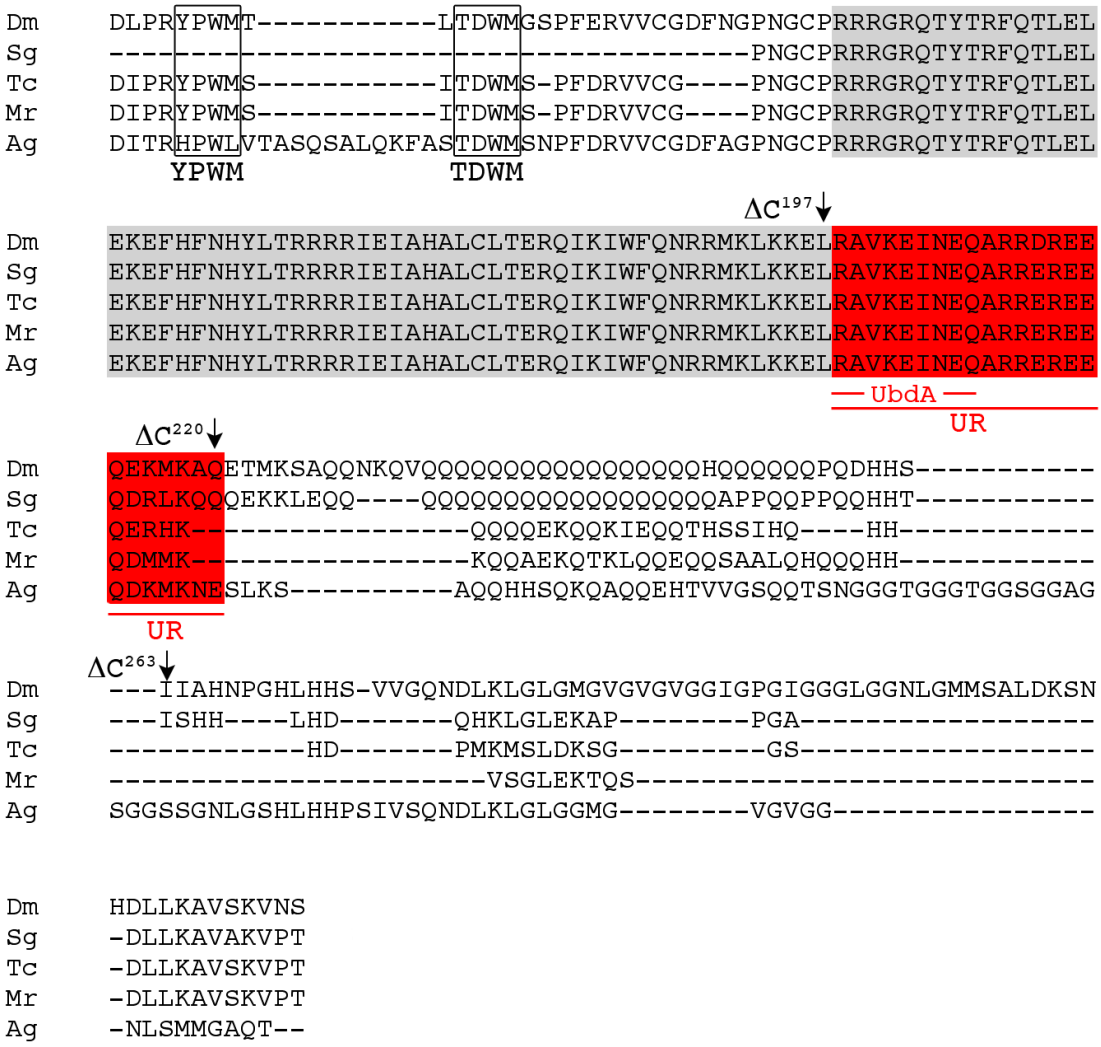


Figure 2.2. Conservation of AbdA motifs.
Sequence alignment (ClustalW) of the predicted amino acid sequences of AbdA orthologs from the arthropods *Drosophila melanogaster* (*Dm*), *Schistocerca gregaria* (*Sg*), *Tribolium castaneum* (*Tc*), *Miyrmica rubra* (*Mr*) and *Anopheles gambiae* (*Ag*). The YPWM and TDWM motifs are boxed, while the HD is in grey. The highly conserved C-terminal UR motif, comprising the previously defined UbdA domain (Merabet et al. 2003) and additional 14 residues, is highlighted in red. The alignment comprises the residues from Asp 105 to Ser 330 of the *melanogaster* ortholog. The N-terminal sequence for *Sg* was not available.

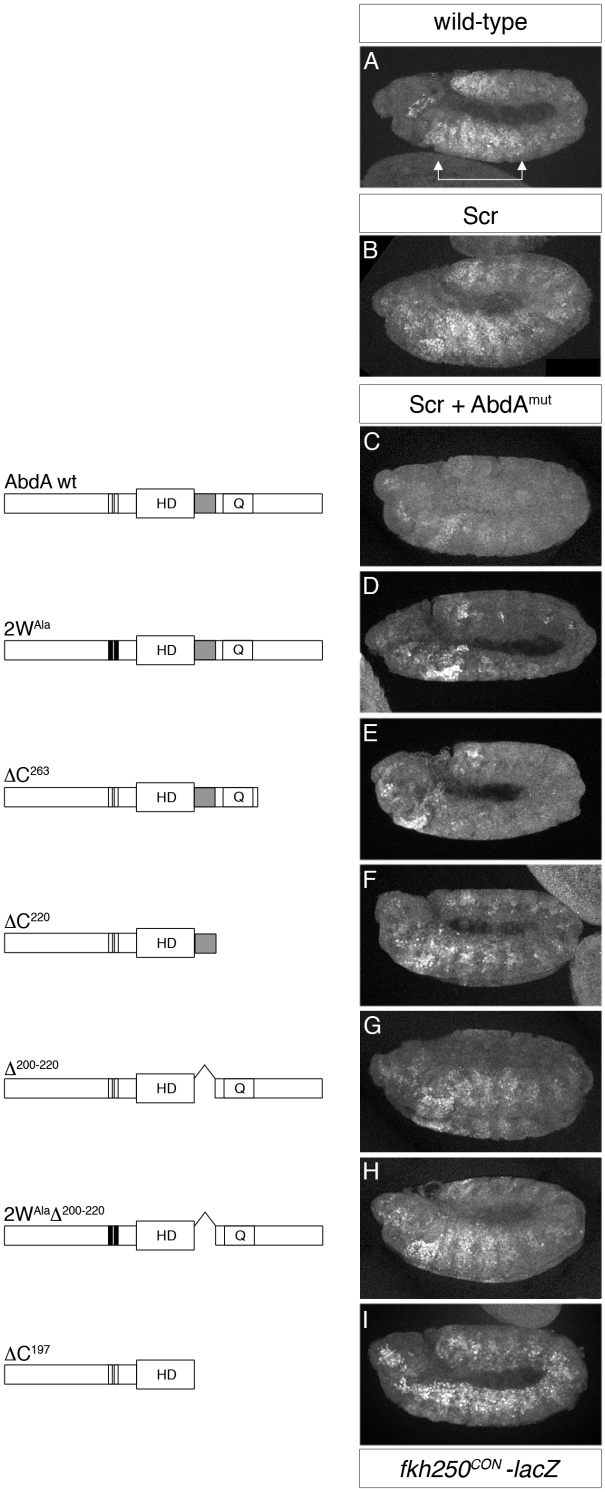


Figure 2.3. **AbdA dominance over Scr depends on its C-terminal UR motif.** AbdA variants (AbdA^{mut}, diagrammed at left) were ectopically expressed in combination with Scr to define motifs in AbdA necessary to suppress Scr activation of *fkh250^{CON}-lacZ* (stained for β-gal, right). Shown are representative images for each AbdA variant, which are ordered according to their ability to repress *fkh250^{CON}-lacZ* from strongest (C, wild-type AbdA) to weakest (I, no C-terminus: ΔC^{197}). Wild-type repressive activity was observed for an AbdA variant in which the YPWM and TDWM motifs were mutated to alanines (D, 2W^{Ala}), suggesting that these motifs are not necessary for repression of this target. Deletion of the entire C-tail of AbdA (I, ΔC^{197}) abolishes repression of *fkh250^{CON}-lacZ*. Repression is significantly rescued by the addition of the UR motif adjacent to the homeodomain, HD (F, ΔC^{220}). Consistently, an internal deletion of UR (G, Δ$^{200-220}$) partially impairs AbdA's ability to suppress activation of *fkh250^{CON}-lacZ* by Scr. A variant in which both the YPWM and TDWM motifs are mutated in combination with this internal deletion (H, 2W^{Ala}Δ$^{200-220}$) displayed no additional loss of repressive ability, suggesting that the UR motif of AbdA is critical for posterior dominance. In the schematics of the AbdA variants, the N-terminal YPWM and TDWM motifs are indicated by white bars (colored black when mutated to alanines) and the C-terminal UR motif is in grey.

Immediately C-terminal to its homeodomain, AbdA contains a so-called UbdA motif, a nine amino acid sequence also present in Ubx, which has been suggested to mediate cooperative binding with Exd to some DNA sequences (Merabet et al., 2007). In fact, UbdA is part of a larger 23 residue conserved region adjacent to the AbdA homeodomain, which we refer to here as the UR motif (for UbdA-RRDR; Figure 2.2). To determine whether this or other regions in the C-tail of AbdA are involved in mediating phenotypic suppression, we tested a series of C-terminal truncations for their ability to compete with Scr for the repression of *fkh250^{CON}-lacZ in vivo*. All AbdA variants were epitope-tagged, allowing us to use transgenes that express at similar levels (Figure 2.4).

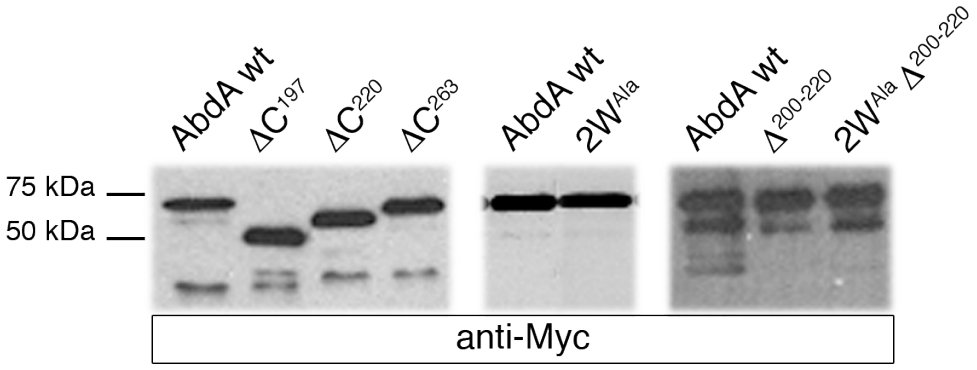


Figure 2.4. AbdA variants are ectopically expressed at similar levels.
Immunoblot of embryonic extracts mis-expressing the indicated AbdA variants stained with an anti-Myc antibody against the N-terminal 6XMyc tag present in all AbdA transgenes used in this study. Equivalent amounts of total embryonic extracts were loaded in each lane, as determined by Bradford assays.

AbdA's ability to compete with Scr for *fkh250^{CON}* regulation is eliminated when the entire C-terminus is removed (Figure 2.3I, ΔC^{197}). Adding back only the UR motif partially restores AbdA's ability to dominate over Scr (Figure 2.3F, ΔC^{220}). Consistently, an internal deletion that removes most of the UR motif (Figure 2.3G, $\Delta^{200\text{-}220}$) exhibits a reduced ability to repress *fkh250^{CON}-lacZ*. No additional loss of repressive activity is

displayed by an AbdA variant in which both the YPWM and TDWM motifs are mutated in combination with this internal deletion (Figure 2.3H, $2W^{Ala}\Delta^{200\text{-}220}$). Additional sequences in the C-tail of AbdA may account for the residual activity of variants lacking the UR motif (Figure 2.3G, H; $\Delta^{200\text{-}220}$ and $2W^{Ala}\Delta^{200\text{-}220}$). All AbdA variants used in this study are capable of repressing the *exd*-independent target gene *spalt* in the wing imaginal disc (Figure 2.5), confirming that these mutants are still functional transcription factors. Furthermore, these mutants retain the ability to repress gene expression *in vivo* arguing that AbdA's repressive activity is not sufficient to account for its ability to dominate Scr. Together, these data highlight the importance of the UR motif for phenotypic suppression.

The above data show that the UR motif is required for AbdA to compete with Scr *in vivo*. To test the hypothesis that UR carries out this function by facilitating cooperative DNA binding with Exd, we analyzed the ability of the truncated AbdA variants to bind *fkh250*^{CON} in complex with Exd (Figure 2.6). In general, the results correlate with the *in vivo* phenotypic suppression assay: those mutants that failed to suppress Scr's ability to activate *fkh250*^{CON}*-lacZ* (ΔC^{197}, $\Delta^{200\text{-}220}$ and $2W^{Ala}\Delta^{200\text{-}220}$) were severely compromised in binding *fkh250*^{CON} with Exd *in vitro* (Figure 2.6). Together, these data strongly suggest that cooperative DNA binding with Exd is required for phenotypic suppression and that domains unique to AbdA are critical for its ability to dominate over Scr. More specifically, they argue that AbdA's UR motif is necessary for cooperative binding of AbdA and Exd to *fkh250*^{CON}, and that the YPWM and TDWM motifs are not sufficient to mediate this interaction on this binding site. The insufficiency of the YPWM motifs to mediate cooperative binding with Exd has been observed for other Hox proteins (Galant et al.,

2002; Merabet et al., 2003; Merabet et al., 2007; Saadaoui et al., 2011), suggesting that the use of paralog-specific motifs such as UR may be a general phenomenon.

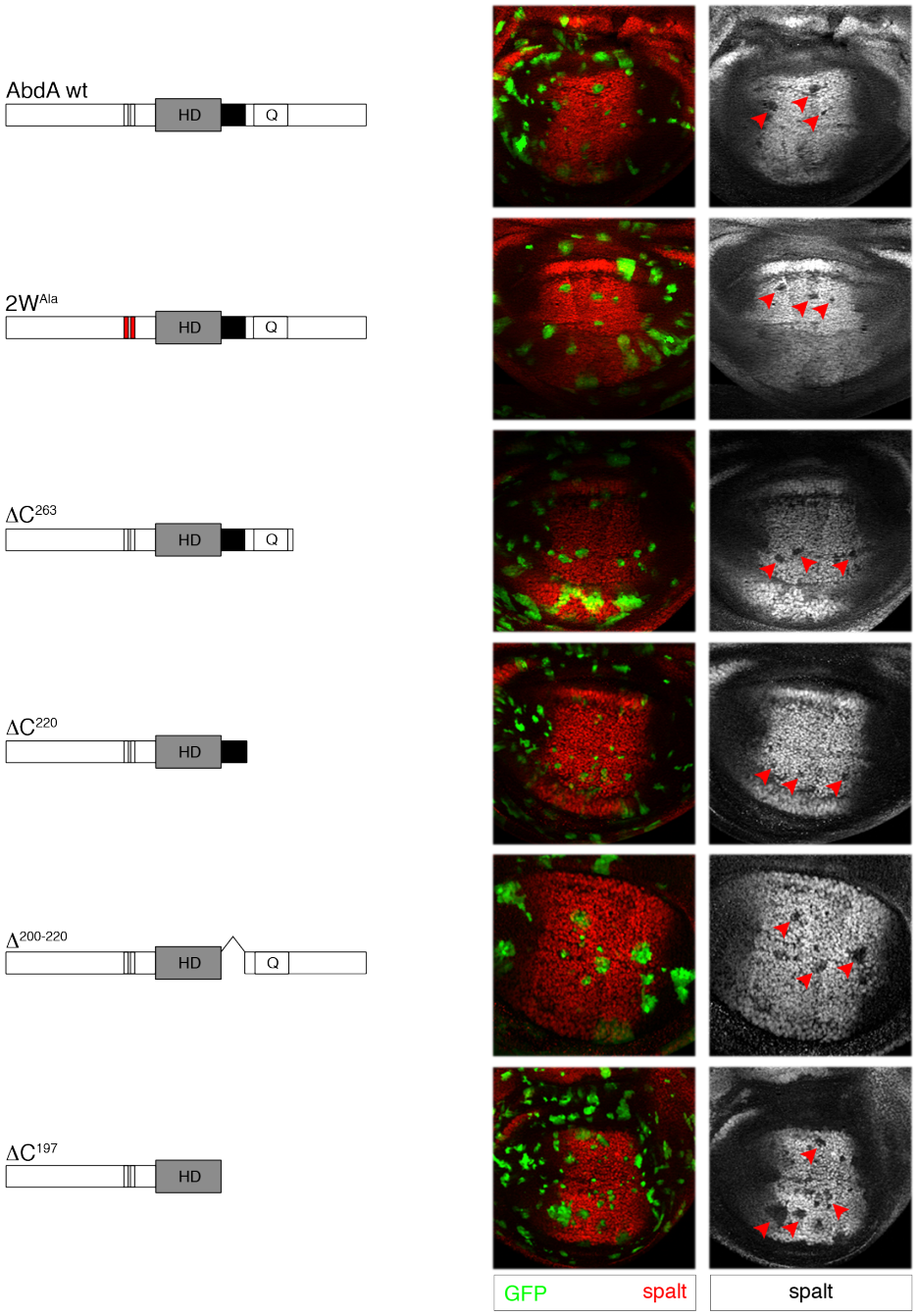


Figure 2.5. AbdA variants repress the *exd*-independent target gene *spalt*. Wing imaginal discs with clones (marked by the presence of GFP in green) ectopically expressing the AbdA variants diagrammed on the left were stained for *spalt* (mid panels: red, right panels: grey). *spalt* was repressed by all AbdA variants tested (except for $2W^{Ala}\Delta^{200\text{-}220}$; no clones could be obtained with this transgene). Red arrowheads point to representative clones.

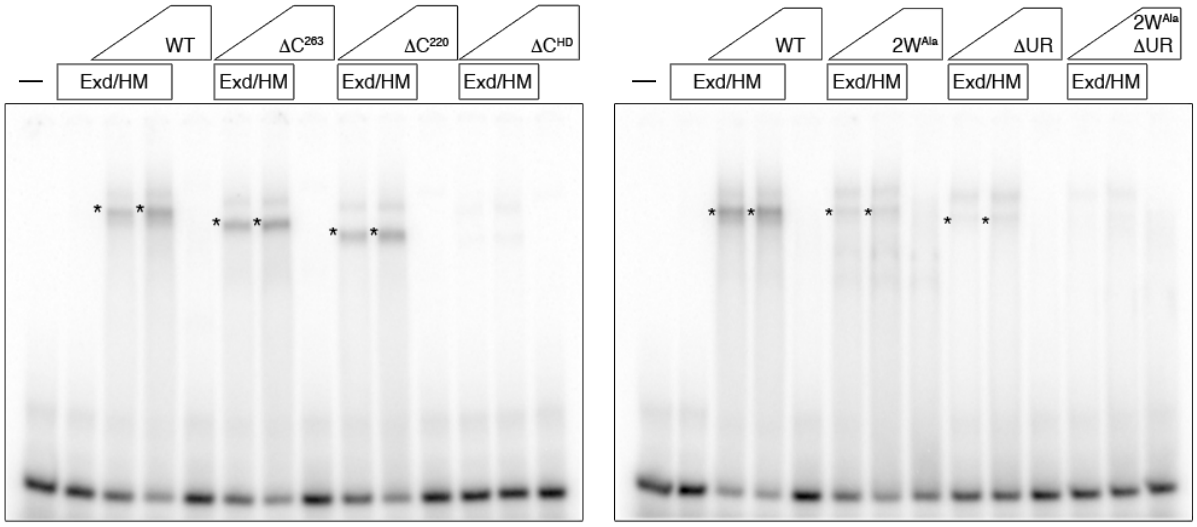


Figure 2.6. The C-terminal UR motif of AbdA mediates cooperative binding with Exd on *fkh250^{CON}*.
Variants of AbdA missing the UR motif (ΔC^{197}, Δ$^{200\text{-}220}$ and 2W^{Ala}Δ$^{200\text{-}220}$) do not form cooperative complexes with Exd on *fkh250^{CON} in vitro*, as analyzed by electrophoretic mobility shift assays (EMSAs). Cooperative complexes between AbdA variants and Exd are indicated by asterisks. EMSAs were performed in the presence of Exd-Hth^{HM} as indicated (see Materials and Methods).

To test the generality of AbdA's dependency on its UR motif for posterior dominance, we analyzed the same AbdA variants for their ability to suppress the activity of the thoracic Hox protein Antp in the patterning of the larval epidermis. When ectopically expressed, Antp transforms the head and first thoracic segment (T1) towards the identity of the second thoracic segment (T2), where Antp is normally expressed (Figure 2.7A). In contrast, when AbdA is ectopically expressed, the head and thorax acquire abdominal segmental identities (Figure 2.7A). Consistent with the rules of phenotypic suppression, wild-type AbdA is able to produce this transformation even in the presence of exogenous Antp (Figure 2.7B). However, similar to the results with *fkh250^{CON}-lacZ*, AbdA mutants that are compromised in their ability to cooperatively bind

DNA with Exd (e.g. ΔC^{197}, $\Delta^{200\text{-}220}$ and $2W^{Ala}\Delta^{200\text{-}220}$) fail to suppress the activity of Antp (Figure 2.7B).

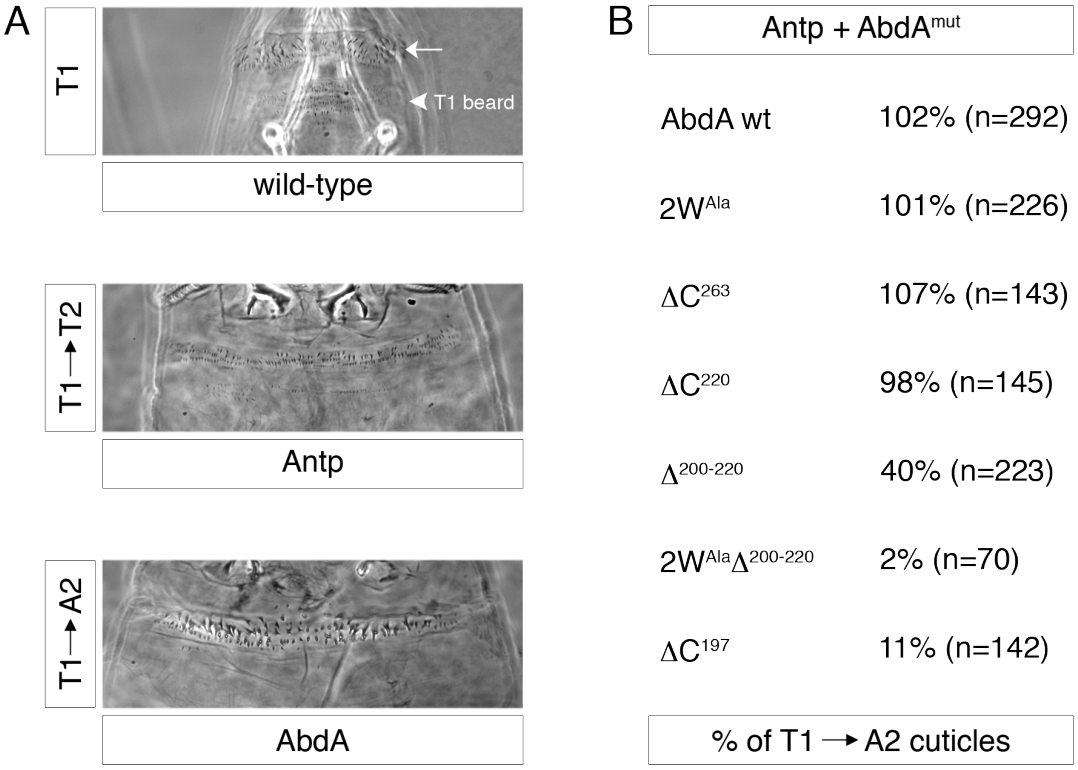


Figure 2.7. **The UR motif endows AbdA with the ability to dominate over Antp in patterning of the larval epidermis.**
Ectopic expression of Antp or AbdA transforms the T1 segment towards a T2 (T1 -> T2) or an A2 fate (T1 -> A2), respectively. Same magnification representative images of wild-type and transformed T1 segments are shown (A). When mis-expressed in combination, wild-type AbdA dominates over Antp generating an A2-like epidermal phenotype (B). AbdA variants ($AbdA^{mut}$) missing the UR motif (ΔC^{197}, $\Delta^{200\text{-}220}$ and $2W^{Ala}\Delta^{200\text{-}220}$) resulted in T2-like phenotypes, suggesting that they were unable to overcome the activity of Antp. No such defect in dominance was observed when only the YPWM and TDWM motifs were mutated ($2W^{Ala}$), suggesting that the UR motif is used by AbdA to dominate over Antp as well as Scr. The percentages of T1 -> A2 cuticles are normalized by the fraction of AbdA-containing embryos, according to the specific genotypes (see Materials and Methods).

Taken together, these data support a model in which phenotypic suppression depends on a competition for cofactor-dependent DNA binding. It follows that this mechanism would only apply to readouts that depend on regulatory elements that are

targeted by multiple Hox proteins. For example, ectopic Scr can activate *fkh* and other target genes required for salivary gland development in more posterior segments, illustrating that this Hox-specific function does not obey phenotypic suppression (Gibson et al., 1990; Jegalian and De Robertis, 1992; Kuziora and McGinnis, 1988). Further, it is particularly noteworthy that, compared to the anterior Hox protein Scr, AbdA has additional motifs that facilitate complex formation with Exd on DNA. Our data suggest that when phenotypic suppression is observed, the more posterior Hox proteins may have a higher affinity for shared binding sites; this higher affinity is a consequence of the quantity and quality of motifs that mediate cooperative DNA binding with Exd. We speculate that these motifs may be used differently at different target genes and binding sites. We suggest that the YPWM motif provides a common, basal level of interaction between Hox proteins and Exd. In the context of Hox-specific regulatory elements, this motif may be sufficient to enable Hox-Exd regulation of some target genes (Joshi et al., 2010). In contrast, in the context of shared enhancers and when multiple Hox proteins are co-expressed, additional, paralog-specific motifs present in the more posterior Hox proteins enable tighter binding of Hox-Exd dimers to DNA, leading to more posterior phenotypes. We have shown this to be the case for a single shared Hox-Exd enhancer and suggest that the generality of this mechanism for phenotypic suppression will become apparent as more shared and specific targets for Hox proteins are characterized at high resolution.

MATERIALS AND METHODS

Fly strains and genetic manipulations

The GAL4/UAS system (Brand and Perrimon, 1993) was used to ectopically express UAS-3XHA-Scr (Joshi et al., 2007) and UAS-6XMyc-AbdA variants in developing embryos. AbdA variants were cloned into p131 and transgenic lines were generated using standard P-element transformation (Abu-Shaar and Mann, 1998). Lines were selected for similar expression levels by western blot (Figure 2.4, mouse anti-Myc 1:1000). The ubiquitous driver *AG11GAL4* was used for all assays. Crosses were grown at 28°C. Embryos were collected every 12 hours and fixed as previously described (Noro et al., 2006). *fkh250-lacZ* and *fkh250^{CON}-lacZ* on the third chromosome were used (Ryoo and Mann, 1999). Flip-out clones were generated by crossing *hs-flp; act<y<Gal4, UAS-GFP* to different UAS-6xMyc-AbdA lines and heat shocking larvae for 10 minutes at 37^{o}C. Wing discs were dissected at wandering stage.

Cuticle preparation

Cuticle preparations were performed as described in Wieschaus and Nusslein-Volhard (Wieschaus, 1986). AbdA variants and Antp were ubiquitously expressed using the *AG11GAL4* and crosses were grown at 28 °C. Homeotic transformations were scored as either “AbdA-like” (head and thoracic segments transformed towards an A2 pattern) or “Antp-like” (head and T1 segments transformed towards a T2 identity). The percentages of T1 -> A2 cuticles are given as the number of embryos with A2-like transformations over the expected number of embryos mis-expressing AbdA, which varies depending on the genotype.

Immunohistochemistry

Rabbit anti-β-galactosidase (Cappell, 1:5000), mouse anti-Myc (Santa Cruz, 1:200) and rat anti-HA (Sigma, 1:500) were used for staining embryos as previously described (Noro et al., 2006). Anti-spalt (Xie et al., 2007), rabbit 1:500) was used for imaginal discs immunostaining. Secondary antibodies used were AlexaFluor488 (1:500), AlexaFluor555 (1:1,000) and AlexaFluor647 (1:500) conjugates from Molecular Probes and embryos were mounted in Vectashield medium. Z-series were collected on a Leica SP5 confocal microscope. All embryonic images shown are Z-projections of the acquired Z-series.

Protein purification and Electrophoretic Mobility Shift Assays (EMSA)

EMSAs were carried out as previously described (Gebelein et al., 2002). The wild-type AbdA construct includes the residues from Ser 79 to the end of the protein (Ryoo and Mann, 1999). Schematics of the AbdA mutants used can be found in Figure 2.3. A nearly full-length wild-type Scr (residues 2-406, Joshi et al., 2007) was used for the EMSAs in Figure 2.1. Since regulation of *fkh250-lacZ* depends on the presence of the HM domain of Hth (Hth^{HM}) to shuttle Exd into the nucleus but does not require the Hth homeodomain (Abu-Shaar et al., 1999; Noro et al., 2006; Stevens and Mann, 2007), Exd-Hth^{HM} heterodimers were co-expressed and purified from *E. coli* and used for all the EMSAs. Hox proteins and Exd-Hth^{HM} were purified as His-tagged fusion proteins using Ni-chromatography (Gebelein et al., 2002). Protein concentrations were determined by the Bradford assay and then confirmed by SDS PAGE and Blue Coomassie analysis (GelCode Blue, Pierce). For EMSAs the following amounts of

proteins were used: Exd-Hth^{HM} 50 ng/lane and AbdA 15 and 45 ng/lane. All the AbdA variants were used in equimolar amounts. The *fkh250^{CON}* probe was previously described (Ryoo and Mann, 1999).

CHAPTER 3.

Variable motif utilization in Hox-cofactor complex formation controls specificity.

Lelli KM, Noro B, and Mann RS. *PNAS.* December 27, 2011 vol. 108 no. 52 21122-21127

This project was started by Barbara Noro as a structure-function analysis of AbdA. As previously mentioned in Chapter 2, she constructed and characterized all of the AbdA mutants except $\Delta^{200\text{-}220}$ and $2W^{Ala}$ $\Delta^{200\text{-}220}$. Based on her observations I constructed and characterized several Scr and Ubx mutants. Although I collected all of the data presented here (with the exception Figure 3.5 panels M, N, O, P, Q and T) Barbara conducted many experiments to characterize several of the AbdA mutants, as well as identifying the UbdA motif as being necessary and sufficient for Engrailed interaction. I wrote the manuscript with helpful comments from Richard Mann and Barbara Noro.

ABSTRACT

Hox proteins often bind DNA cooperatively with cofactors such as Extradenticle (Exd) and Homothorax (Hth) to achieve functional specificity *in vivo*. Previous studies identified the Hox YPWM motif as an important Exd interaction motif. Using a comparative approach, we characterize the contribution of additional conserved sequence motifs to the regulation of specific target genes for three *Drosophila* Hox proteins. We find that Sex combs reduced (Scr) uses a simple interaction mechanism, where a single tryptophan-containing motif is necessary for Exd-dependent DNA-binding and *in vivo* functions. Abdominal-A (AbdA) is more complex, using multiple conserved motifs in a context dependent manner. Lastly, Ultrabithorax (Ubx) is the most flexible, in that it uses multiple conserved motifs that function in parallel to regulate target genes *in vivo*. We propose that using different binding mechanisms with the same cofactor may be one strategy to achieve functional specificity *in vivo*.

INTRODUCTION

Understanding the molecular processes by which gene expression is regulated remains at the core of many biological questions. The predominant model of eukaryotic gene regulation emphasizes the role of site-specific transcription factors in target gene selection. The initial binding of these transcription factors anchors the rest of the transcriptional regulatory complex, or enhanceosome, to the target site. Recruitment of additional proteins is often required to determine the regulatory sign, whether a gene is activated or repressed, and if this state will be maintained. In some cases the DNA sequence can provide considerable insight into which other proteins are recruited

(Gebelein et al., 2004). However, enhanceosome formation also requires protein-protein interactions: mutational analyses of transcription factors demonstrate that sequences outside of the DNA-binding domain can influence regulatory activity, in part by influencing the assembly of DNA-bound protein complexes (Georges et al., 2010; Mann et al., 2009).

The Hox or homeotic selector proteins provide a powerful system in which to study the role of protein-protein interactions in enhanceosome formation and transcription factor function. Best known for their role in anterior-posterior patterning, Hox proteins contain a highly conserved DNA-binding domain, termed the homeodomain (McGinnis and Krumlauf, 1992). Although most homeodomains bind similar AT-rich sequences *in vitro* (Noyes et al., 2008), each Hox protein displays a high level of functional specificity *in vivo* (Hughes and Kaufman, 2002). These observations imply that residues outside of the DNA-binding domain influence specificity *in vivo*. One way Hox proteins achieve higher specificity is through cooperative interactions with DNA-binding cofactors such as Extradenticle (Exd in *Drosophila*; Pbx in vertebrates) and Homothorax (Hth in *Drosophila*; Meis in vertebrates) (Mann and Chan, 1996; Moens and Selleri, 2006). Exd and Hth, both members of the three-amino-acid loop extension (TALE) family of homeodomain proteins, are obligate dimer partners for both nuclear translocation and transcriptional activity *in vivo* (Rieckhof et al., 1997; Stevens and Mann, 2007). Previous genetic analyses highlight the importance of *exd* and *hth* for Hox function during embryogenesis (Noro et al., 2006; Peifer and Wieschaus, 1990; Rieckhof et al., 1997). In addition to Exd and Hth, the abdominal Hox proteins Ultrabithorax (Ubx) and AbdominalA (AbdA) also have the ability to bind cooperatively

with another homeodomain protein, Engrailed (En) (Gebelein et al., 2004). As with Exd-Hth-Hox, En-Hox-DNA complex formation has been shown to be critical for Hox-mediated gene regulation *in vivo* (Gebelein et al., 2004).

Biochemical and X-ray crystallography studies identified the highly conserved Hox motif called YPWM as one mode by which Hox proteins interact with Exd-Hth. However, despite being evolutionarily conserved and present in most Hox proteins, the importance of the YPWM motif appears to vary (Mann et al., 2009). For example, while vertebrate Hoxa1 and Deformed (Dfd) require the YPWM motif for some Pbx/Exd-dependent functions (Delval et al., 2011; Green et al., 1998; Joshi et al., 2010; Remacle et al., 2004), Ubx and AbdA do not require YPWM for some Exd-dependent functions (Galant et al., 2002; Merabet et al., 2003; Merabet et al., 2007; Saadaoui et al., 2011; Tour et al., 2005). In the case of Ubx and AbdA, a shared six amino acid motif C-terminal to the homeodomain, termed UbdA, has been suggested to also contribute to interactions with Exd (Chan et al., 1994; Merabet et al., 2007; Saadaoui et al., 2011). Interestingly, in addition to UbdA, both Ubx and AbdA have other evolutionarily conserved residues in the C-terminus that may also be important for mediating Hox functions *in vivo* (Galant and Carroll, 2002; Hittinger et al., 2005; Merabet et al., 2003; Ronshaugen et al., 2002). These C-terminal sequences, which are distinct in Ubx and AbdA, could modify UbdA–dependent interactions so that its function may not be equivalent in both proteins. Further, the presence of multiple Exd-interaction motifs poses the question of what determines which mode of interaction is most relevant for a given *in vivo* function. In the case of Ubx, both the protein context and the target site

have been suggested to influence how individual motifs are used (Saadaoui et al., 2011).

Currently, it is generally unknown how different modes of cofactor interaction influence Hox specificity. In the case of Sex combs reduced (Scr), structural studies demonstrated that the YPWM-Exd interaction helps position amino acids of the Hox linker region, which separates YPWM from the homeodomain, to make critical contacts in the minor groove of a specific DNA binding site (Joshi et al., 2007). These observations raise the possibility that other modes of Exd-Hox interaction could also alter the way in which these protein complexes recognize and bind to their target DNA sequences. In addition, alternate modes of protein-DNA complex formation may also impact the recruitment of coactivators and corepressors, as has been suggested for the Glucacorticoid receptor (Meijsing et al., 2009).

In the current study, we use a comparative approach to characterize conserved sequence motifs for three Hox proteins and analyze their requirement for different *in vivo* functions. We demonstrate that the YPWM motif is critical for Scr to carry out Exd-dependent functions *in vivo*. In contrast to the single Exd-interaction motif of Scr, AbdA and Ubx are more complex. In addition to the previously described YPWM and UbdA motifs, both AbdA and Ubx have other conserved motifs that contribute to cooperative complex formation with Exd-Hth *in vitro*. However, the *in vivo* requirements for conserved motifs vary according to the Hox protein. AbdA uses motifs in a context dependent manner, whereas Ubx is more flexible, with some motifs apparently acting in a redundant manner for certain readouts. Our results suggest that having multiple

sequence motifs adds complexity to the assembly and function of Hox complexes *in vivo*.

RESULTS

Scr requires its YPWM motif for Exd-dependent functions.

Previous studies demonstrated that the YPWM motif of Scr makes direct contacts within the Exd homeodomain and contributes to Scr function *in vivo* (Joshi et al., 2007; Papadopoulos et al., 2010). Because, as shown below, some Hox proteins have multiple YPWM-like motifs, we rename them here as W motifs and number them according to their proximity to the homeodomain. To specifically test the requirement of Scr's single W motif, we constructed a mutant variant in which this sequence (YPWM) was mutated to four alanines (Scr-W1A; Figure 3.1B). This mutant was assayed both for its ability to cooperatively bind with Exd to DNA *in vitro* and to perform Exd-dependent functions *in vivo*. One well-characterized readout is the Scr-specific target gene *forkhead* (*fkh*), which utilizes an Scr-Exd DNA-binding site, *fkh250*, for *in vivo* regulation (Ryoo and Mann, 1999). While strong cooperative binding was observed for Scr-WT and Exd to *fkh250*, Scr-W1A was unable to bind cooperatively with Exd to this site (Figure 3.1A). In contrast, both Scr-WT and Scr-W1A were able to bind to DNA as monomers *in vitro* (Figure 3.1A), demonstrating that these proteins are still competent to bind DNA.

To test both wild type and mutant proteins for *in vivo* function, we used the Gal4-UAS system to misexpress wild type and mutant versions of Scr *in vivo* (Brand and Perrimon, 1993). In wild type embryos endogenous expression of Scr activates *fkh250*-

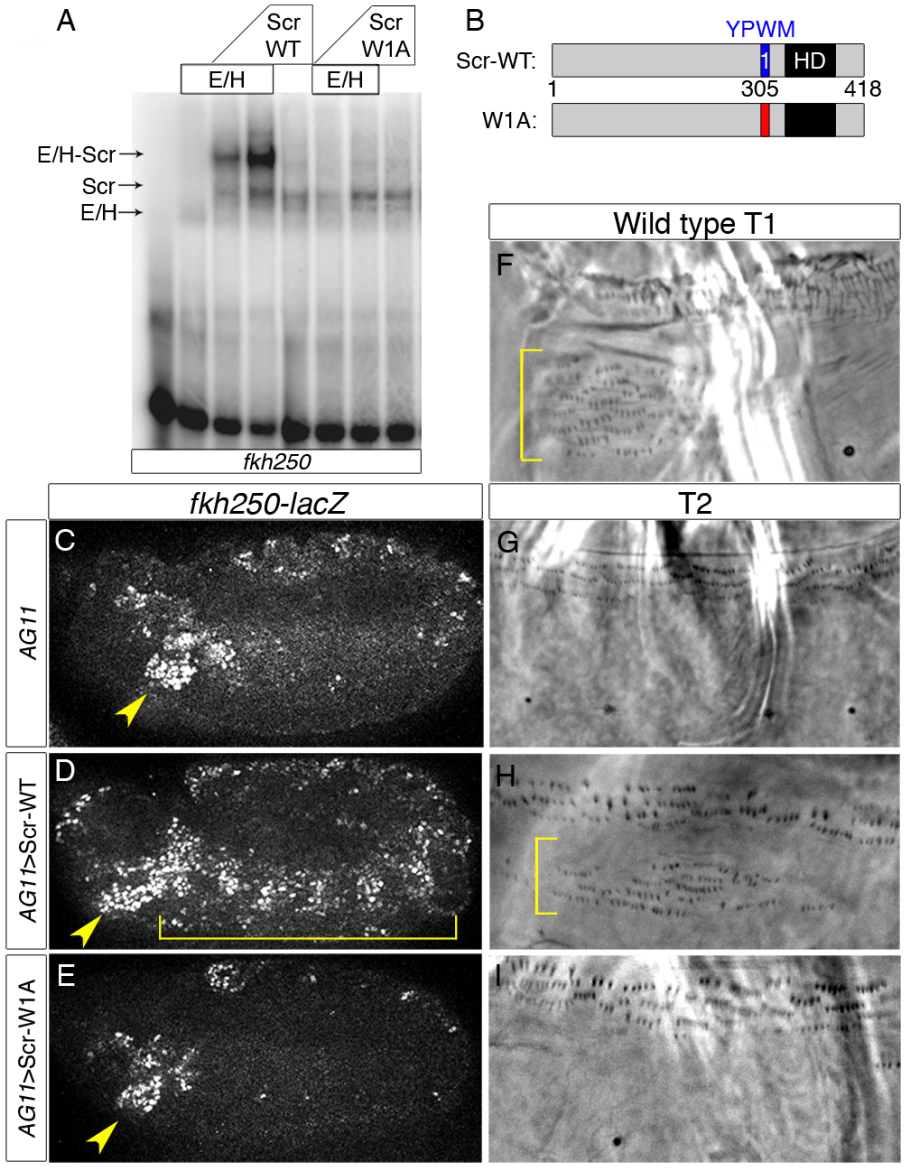


Figure 3.1. Scr requires the W motif for Exd-dependent functions.
(A) EMSA of Scr proteins with Exd-Hth^{HM} on *fkh250*. Hth^{HM} is a homeodomain-less isoform of Hth that is sufficient for *fkh250-lacZ* regulation *in vivo* (Noro et al., 2006). Complexes are indicated with arrows. (B) Schematics of wild type and mutant Scr proteins. Homeodomain (HD) in black. Blue designates the W-motif. Red indicates residues mutated to alanines (YPWM->AAAA).(C-E) *fkh250-lacZ* expression. *AG11-Gal4* is a ubiquitous Gal4 driver controlled by *armadillo*. Embryos were stained for ßgalactosidase (ßgal) to monitor *fkh250-lacZ* expression (white). Yellow arrowheads indicate areas of wild type *lacZ* expression. Brackets indicate induction of ectopic *lacZ* expression. (F) Phase contrast image of a wild type T1 ventral denticle pattern normally controlled by Scr. (G-I) Phase contrast images depicting ventral T2 cuticle patterns for wild type larvae (G) or animals ectopically expressing wild type Scr (H) or Scr-W1A (I) using the *AG11-Gal4* driver.

lacZ only in parasegment 2 (Figure 3.1C). When Scr was ubiquitously expressed throughout the embryo ectopic activation of the *fkh250-lacZ* reporter gene was observed (Figure 3.1D). In contrast, Scr-W1A was unable to activate this reporter (Figure 3.1E). Scr-W1A was also analyzed for its ability to induce homeotic transformations of the larval cuticle. When wild type Scr was ubiquitously expressed throughout the embryo, ectopic hairs similar to those found in the first thoracic (T1) segment (called the T1 beard, Figure 1F) were observed in additional segments (Figure 3.1H compared to G). In contrast, ubiquitous expression of Scr-W1A did not induce any ectopic beard (Figure 3.1I). However, both Scr-WT and Scr-W1A were able to repress the Exd-independent target *spalt (sal)* in the wing imaginal disc, demonstrating that Scr-W1A is still an active transcription factor (Figure 3.2). From these results we conclude that Scr's W motif is crucial to form cooperative complexes with Exd *in vitro* and execute Exd-dependent, but not Exd-independent, functions *in vivo*. These observations are consistent with the observation that severely truncated forms of Scr that retain its W motif and homeodomain retain the ability to generate Scr-like transformations *in vivo* (Papadopoulos et al., 2010).

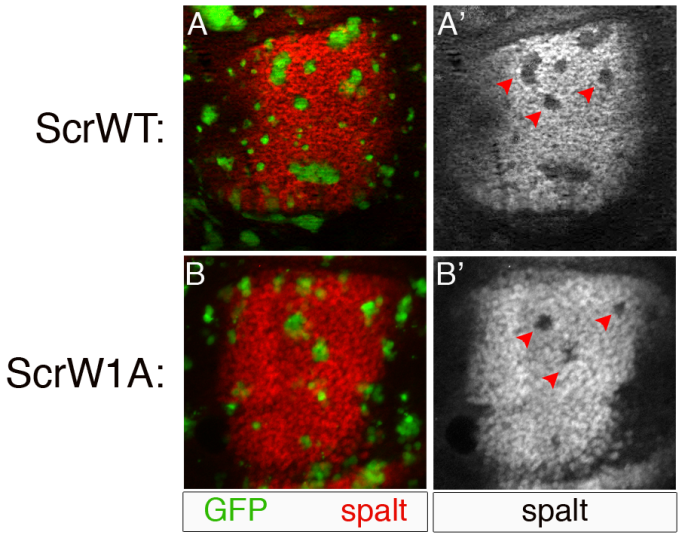


Figure 3.2. Scr proteins repress *sal* in the wing imaginal disc.
(A,B) Pouch region of wing imaginal discs containing clones ectopically expressing Scr (A) or Scr-W1A (B) stained for GFP (which marks the clones) and Sal (red or white). Representative clones are indicated with red arrowheads.

Multiple conserved motifs contribute to Exd-dependent functions of AbdA in a context dependent manner.

In addition to its classical YPWM motif (AbdA-W2, Figure 2A), AbdA has another putative cofactor interaction motif C-terminal to its homeodomain, UbdA (AbdA-U, Figure 2A) (Chan and Mann, 1993). Based on sequence conservation among arthropods we identified two additional sequence motifs: the W motif TDWM (AbdA-W1) and RRDR (AbdA-R), C-terminal to UbdA (Figure 3.3). To test the role of each motif, both individually and in combination, we made a series of AbdA mutants (Figure 3.4A). In the abdomen of the developing embryo, AbdA normally represses the limb-forming gene *Distalless* (*Dll*) to restrict leg development to the thorax (Vachon et al., 1992). Previous studies characterized the DNA binding sites abdominal Hox proteins and their cofactors use to directly repress *Dll* in the abdomen (Gebelein et al., 2002; Gebelein et al., 2004). We analyzed the contribution of the conserved motifs of AbdA for cooperative complex formation with Exd/Hth on the *DMX-R1* binding site (Figure 3.4B). All of these motifs individually, except for W1, contribute to AbdA's ability to form complexes with Exd-Hth on the *DMX-R1* binding site (Figure 3.4B). Moreover, we often observed an additional reduction in complex formation when multiple motifs were mutated, suggesting that they contribute in an additive manner to complex formation with Exd-Hth. For example, the W1&2A;RA mutant (in which W1, W2, and RRDR motifs are mutated) was significantly impaired in complex formation compared to the W1&2A or the RA mutants individually (Figure 3.4B). In contrast, when binding was assayed on a different site (*rhoA*) where the functional interaction is between AbdA and Hth (Li-Kroeger et al., 2008; Uhl et al., 2010), complex formation was only minimally affected,

suggesting that these motifs contribute specifically to Exd-dependent DNA binding (Figure 3.4C).

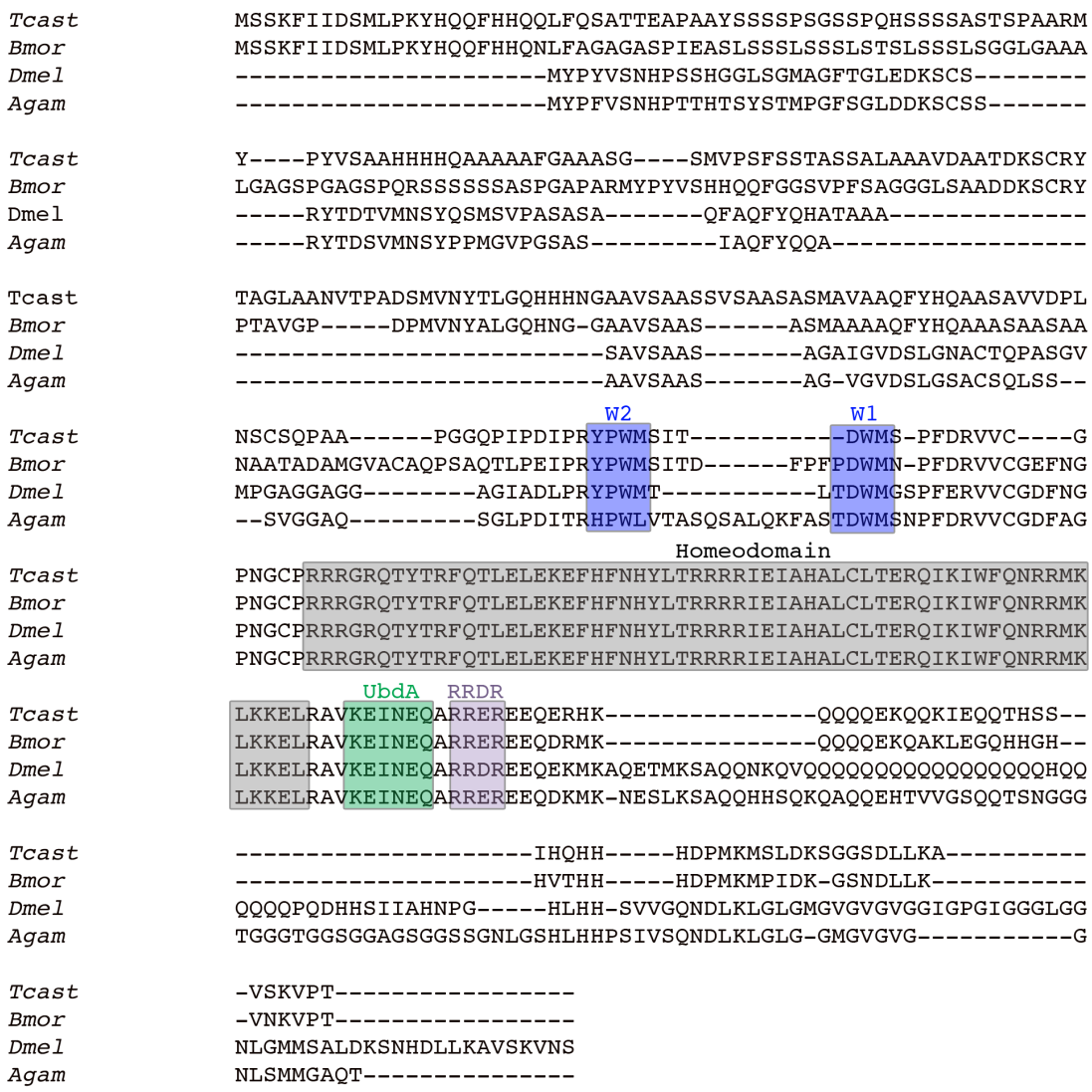


Figure 3.3. AbdA motifs are conserved.
Alignment of insect AbdA proteins: *Bmor* (*Bombyx mori*), *Tcast* (*Tribolium castaneum*), *Dmel* (*Drosophila melanogaster*), *Agam* (*Anopheles gambiae*). Relevant motifs are highlighted. Alignment was done using default settings of ClustalW.

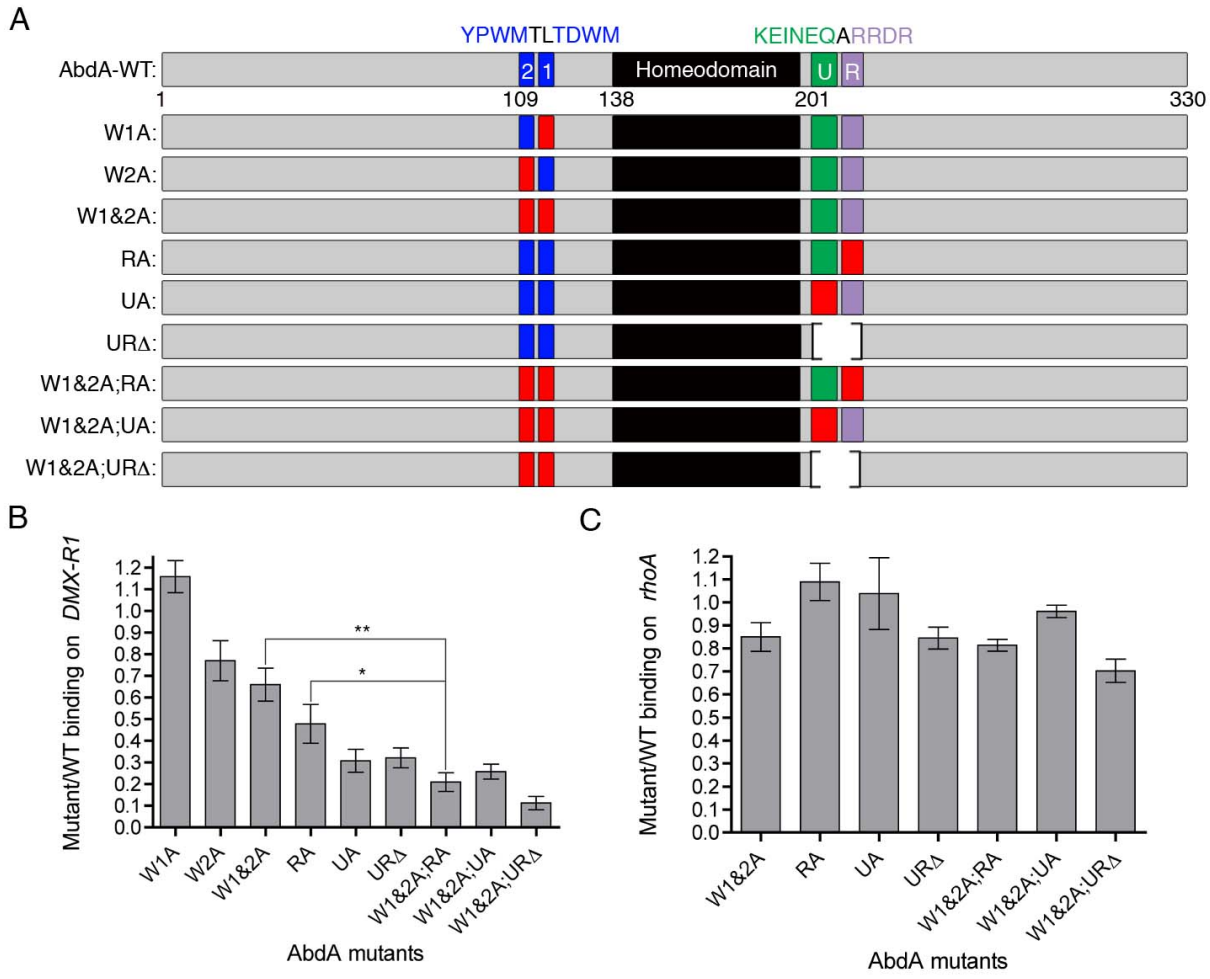


Figure 3.4. AbdA has multiple motifs that mediate cooperative complex formation with Exd/Hth.

(A) Schematics of wild type and mutant AbdA proteins. Diagrams are approximately to scale. Homeodomain (HD) in black. Blue designates W-motifs. Green designates the UbdA motif (U). Purple designates the RRDR motif (R). Red indicates residues mutated to alanines (YPWM->AAAA, TDWM->AAAA, KEINEQ->AAAAAA) and RRDR->AAAA). (B) Average binding of different AbdA mutants to the *DMX-R1* probe. T-tests were used to determine if the difference in cooperative binding is significant for a subset of mutants (** p-value=0.004 and * p-value=0.031). (C) Average binding of different AbdA mutants with Hth-Exd to the *rhoA* probe. The amount of cooperative complex formation as a ratio of wild type binding did not change significantly across the different mutants (ANOVA analysis: $F(6,13)=2.664$ p-value=0.065). Bar graph represents the mean of $n \geq 3$ ratios from independent experiments for each mutant. Error bars represent standard error of the mean (SEM).

We next compared the ability of each mutant to repress *Dll in vivo* using *prdGal4* to drive expression of AbdA variants in the second thoracic segment (T2). Wild type *Dll* expression in the first and third thoracic segments, where *prdGal4* is not expressed, was used as reference (T1 and T3 respectively, Figure 3.5A-K). These data parallel the *in vitro* results: most motifs contribute to *Dll* repression to some extent and often show additive affects when tested in combination. For example, additional loss of *Dll* repression was observed for the W1&2A mutant compared to the two single mutants (Figure 3.5E compared to C and D), as well as for the W1&2A;RA mutant compared to W1&2A and RA (Figure 3.5I compared to E and F). Unlike the W motifs, UbdA is necessary for all of the *in vivo* functions of AbdA analyzed (Figure 3.5G and R). For example, mutation of UbdA (UA) alone abolished AbdA's ability to repress *Dll* (Figure 3.5G). Therefore, no additive effects were observed in compound mutants containing mutations in UbdA (Figure 3.5H, J and K). However, although necessary, UbdA is not sufficient to impart wild type repressive ability: the AbdA-W1&2A;RA mutant, in which UbdA is still intact, did not repress *Dll* (Figure 3.5I). As with the Scr-W1A mutant, all of the AbdA mutants were able to repress *sal* in the wing imaginal disc (Figure 3.6), demonstrating that these proteins are still active transcription factors and that the motifs mapped here are required for Exd-dependent functions *in vivo*.

AbdA also patterns the denticle belts in the second through eighth abdominal segments (A2-A8) of the *Drosophila* larva (wild type A2 shown in Figure 3.5W) (Sanchez-Herrero et al., 1985). Ectopic expression of AbdA causes homeotic transformations to A2 in segments anterior to the wild type A2 segment (Figure 3.5M compared to L) (Merabet et al., 2003). By ubiquitously misexpressing AbdA variants

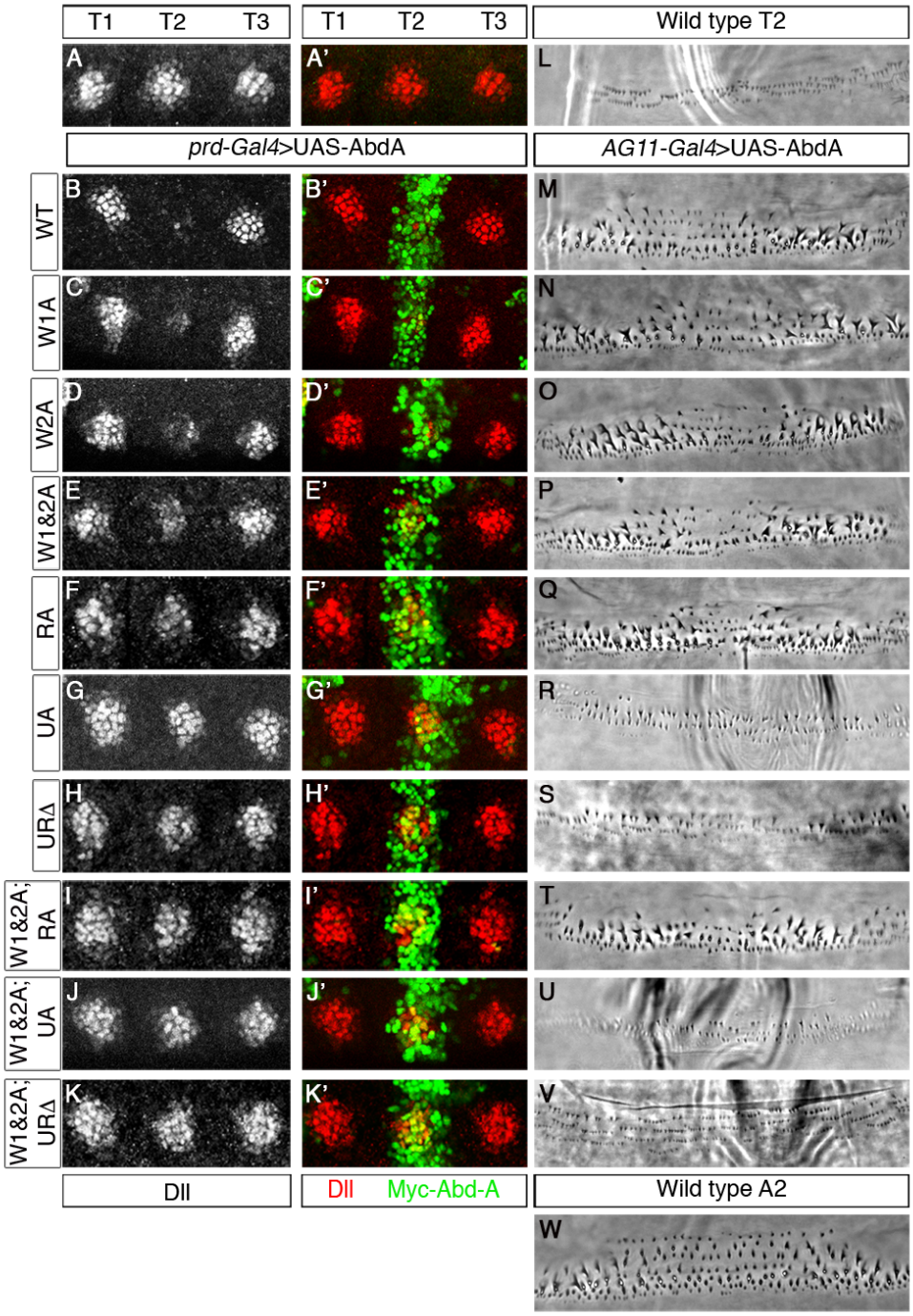


Figure 3.5. AbdA uses conserved motifs in a context dependent manner.
(A-K) Thoracic region of a wild type embryo (A) or embryos expressing AbdA proteins in T2 via the *prd-Gal4* driver, stained for *Dll* (white or red) and myc-AbdA (green). The protein variant used is indicated on the left. (L-V). Phase contrast images depicting ventral T2 cuticle patterns for wild type larvae (L) or animals ectopically expressing wild type AbdA (M) or mutant variants (N-V) using the *AG11-Gal4* driver. (W) A wild type A2 ventral denticle pattern normally controlled by AbdA is shown.

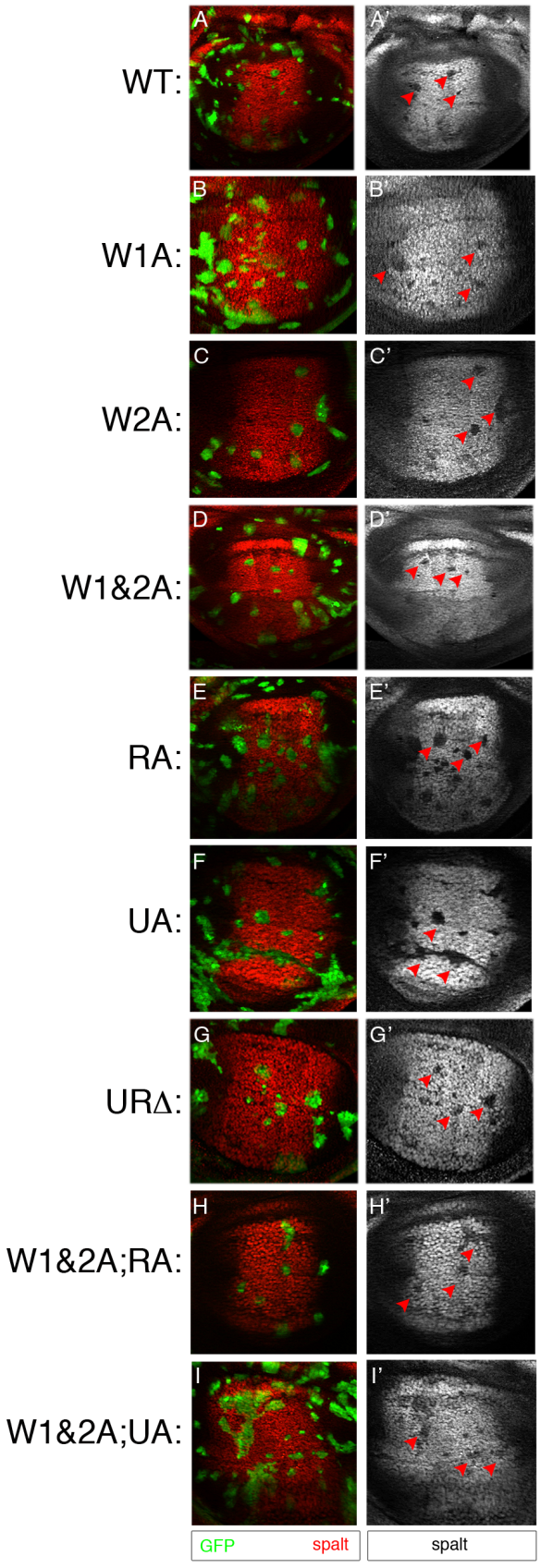


Figure 3.6. AbdA proteins repress *sal* in the wing imaginal disc.
(A-I) Pouch region of wing imaginal discs containing clones ectopically expressing AbdA variants stained for GFP (green, which marks the clones) and Sal (red or white). Representative clones are indicated with red arrowheads.

throughout the embryo we tested the effect different mutations have on the ability to confer segment identity (Figure 3.5M-V). Similar to loss of *Dll* repression, mutation of UbdA alone, or in combination with other motifs, prevented abdominal-like transformations (Figure 3.5R, S, U and V). However, AbdA mutants in which UbdA was intact were able to induce strong abdominal transformations (Figure 3.5N, O, P, Q and T). Comparing the contribution of each motif for different AbdA functions therefore suggests that motif usage is target-dependent. For example, while W1&2A;RA, which has an intact UbdA motif, still induced AbdA-like cuticle transformations it did not repress *Dll* (Figure 3.5I and T).

Ubx has a different requirement for motifs shared with AbdA.

Based on the knowledge gained from the AbdA analyses and sequence conservation (Figure 3.7), we performed a similar structure-function study of another abdominal Hox protein, Ubx, which is a repressor of *Dll* like AbdA. Also like AbdA, Ubx has multiple conserved W motifs and a UbdA motif, as well as a previously identified QA motif in an analogous location to the RRDR motif of AbdA (Galant and Carroll, 2002; Hittinger et al., 2005). However, unlike AbdA, point mutations in Ubx's UbdA motif (to AALVAV) did not hinder cooperative complex formation with Exd-Hth on *DMX-R1 in vitro* or the ability to repress *Dll in vivo* (Figure 3.8A-D). Previous studies have shown that a different set of point mutations in the UbdA motif (to VVLIVA) prevented cooperative complex formation with Exd-Hth on *DMX-R1* (Figure 3.8A) (Merabet et al., 2007). However, we find that this set of point mutations, but not the AALVAV mutant, resulted in a decrease in the ability of Ubx to bind as a monomer (Figure 3.8A and E). Therefore,

any decrease in cooperative complex formation observed with the VVLIVA mutant may not be due to a loss of interaction with Exd/Hth but instead to the compromised ability of this mutant to bind any DNA sequence. Moreover, and in agreement with the AALVAV UbdA mutant, deletion of the entire C-terminus (Ubx-CTΔ), including the UbdA motif, did not affect its ability to repress *Dll* or induce homeotic transformations *in vivo* (Figure 3.9D and P). Thus, while the UbdA motif is critical for AbdA to execute its *in vivo* functions, it is dispensable in Ubx.

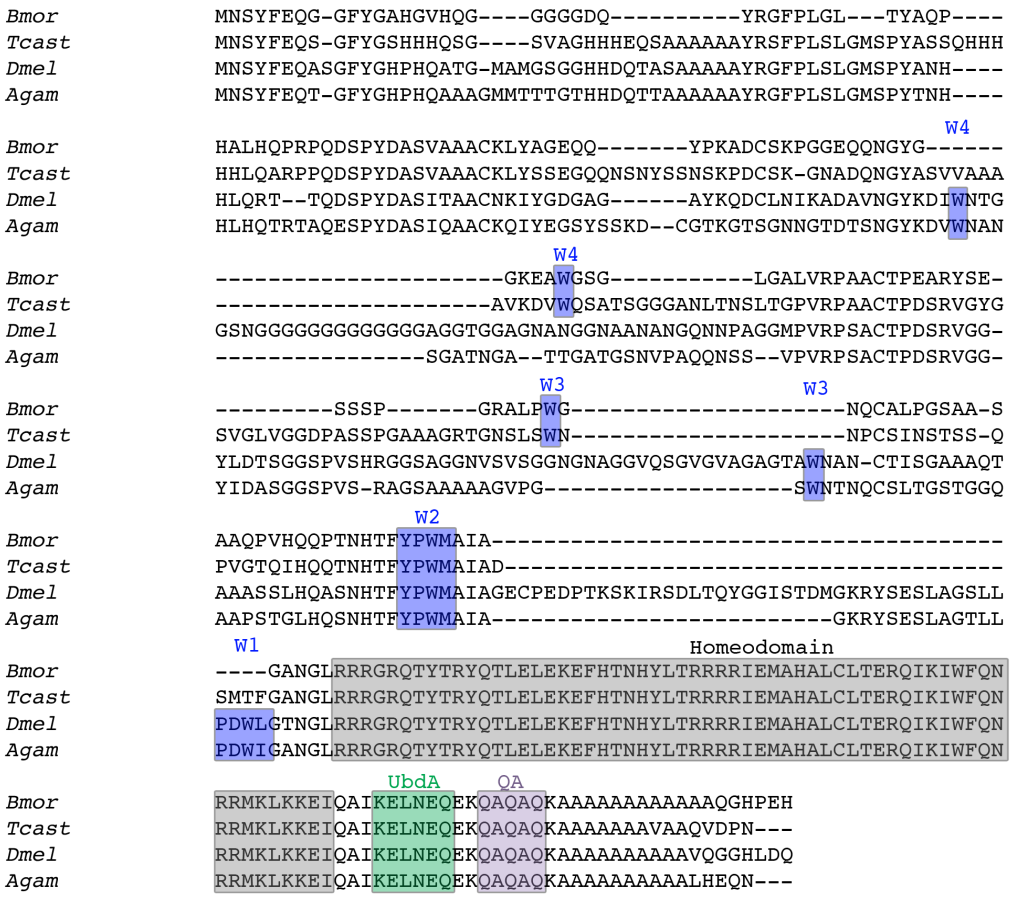


Figure 3.7. Ubx motifs are conserved.
Alignment of insect Ubx proteins: *Bmor* (*Bombyx mori*), *Tcast* (*Tribolium castaneum*), *Dmel* (*Drosophila melanogaster*), *Agam* (*Anopheles gambiae*). The Dmel sequence corresponds to the Ubx1a isoform. Relevant motifs are highlighted. Alignment was done using default settings of ClustalW.

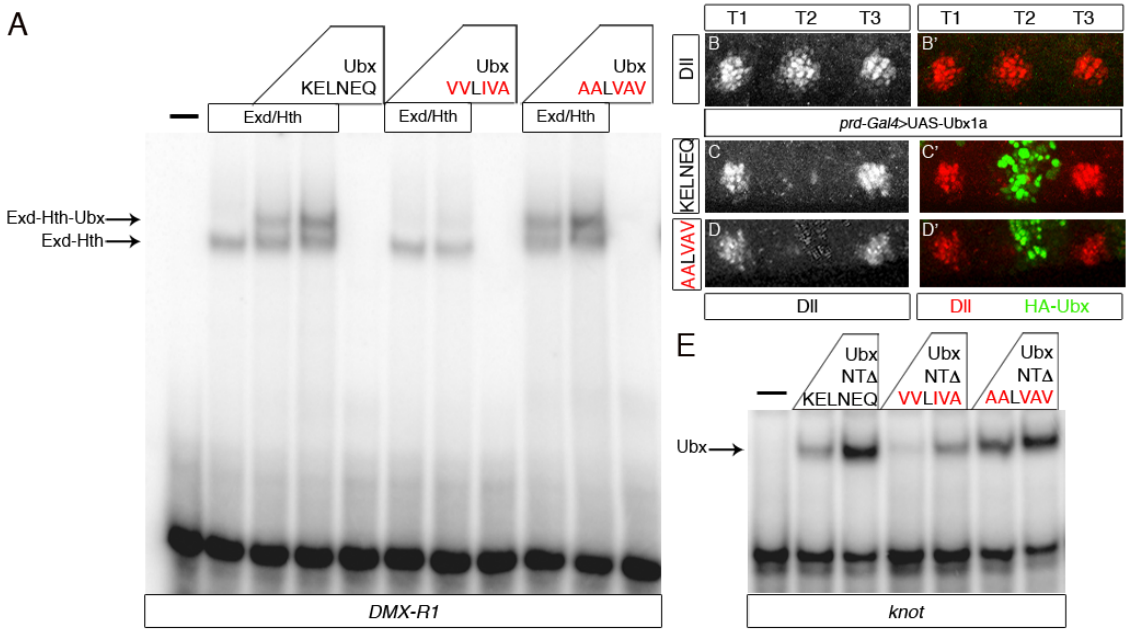


Figure 3.8. Point mutations in UbdA can adversely affect monomer DNA-binding. (A) Binding of Ubx proteins (aa57-380) with Exd-Hth on *DMX-R1*. Point mutations in UbdA are indicated in red. Cooperative complexes are designated by arrows. The wild type UbdA sequence is KELNEQ. When this sequence was mutated to VVLIVA as described by (Merabet et al., 2007), a loss of cooperative binding was observed. However, when the same residues were mutated to AALVAV, complex formation was not compromised. (B-D) Thoracic region of a wild type embryo (B) or embryos expressing Ubx proteins in T2 via the *prd-Gal4* driver, stained for *Dll* (white or red) and HA-Ubx (green). The protein variant is indicated on the left. The AALVAV UbdA mutant is still a potent repressor of *Dll*, while the VVLIVA mutant (Merabet et al., 2007) is not. (E) EMSA of Ubx-NTΔ proteins on the monomer Ubx binding site from *knot* (Hersh et al., 2007). Point mutations in UbdA are indicated in red. Monomer Ubx binding is designated by an arrow. N-terminally truncated proteins were used to assess binding in the absence of any potential inhibitory effects of N-terminal sequences (Chan et al., 1996). The VVLIVA mutant of UbdA hinders monomeric Ubx binding, while the AALVAV mutant does not. From these data, we conclude that the failure of the VVLIVA mutant to repress *Dll* (Merabet et al., 2007) is likely to be a consequence of compromised homeodomain-DNA interaction, and not due to a defect in cooperative complex formation with Exd as suggested previously (Merabet et al., 2007).

Starting with the Ubx-CTΔ mutant, we then mutated each of four potential W motifs (Figure 3.9). Surprisingly, all of these mutants, even when all putative W motifs were mutated in the context of Ubx-CTΔ, were able to repress *Dll* (Figure 3.9D-G). Moreover, all of these Ubx mutants were also able to activate *dpp674-lacZ*, an Exd-dependent Ubx target in the visceral mesoderm (Capovilla et al., 1994; Chan et al., 1994) (Figure 3.9J-M). Ubx also controls the formation of a specific pattern of denticles in the first abdominal segment (A1) of *Drosophila* larvae (Figure 3.9T) (Duncan, 1987). Ectopic expression of Ubx causes homeotic transformations to A1 in segments anterior to the wild type A1 segment (Figure 3.9O compared to N and T) (Gonzalez-Reyes and Morata, 1990; Mann and Hogness, 1990). Remarkably, all mutants in the CTΔ series were also able to transform T2 to an A1-like identity (Figure 3.9P-S). Similar results for *Dll* repression, *dpp674-lacZ* activation and cuticle transformation were obtained when the same series of mutations were tested in the context of a shorter isoform of Ubx, UbxIVa, in which a small portion of the linker region is removed by alternative splicing (Figure 3.10) (O'Connor et al., 1988). Unfortunately, truncation mutants aimed at identifying essential sequences in the N-terminus were uninformative due to a lack of nuclear localization and/or instability of the truncated proteins *in vivo* (Addendum 1). However, it is unlikely that the Ubx N-terminus has additional, non-W motifs for interacting with Exd because mutating the W motifs in Ubx-CTΔ effectively eliminated cooperative complex formation with Exd on the *DMX-R1* binding site (Figures 3.10B and 3.11). Therefore, these results raise the possibility that Ubx may execute these *in vivo* functions at least in part in an Exd-independent manner.

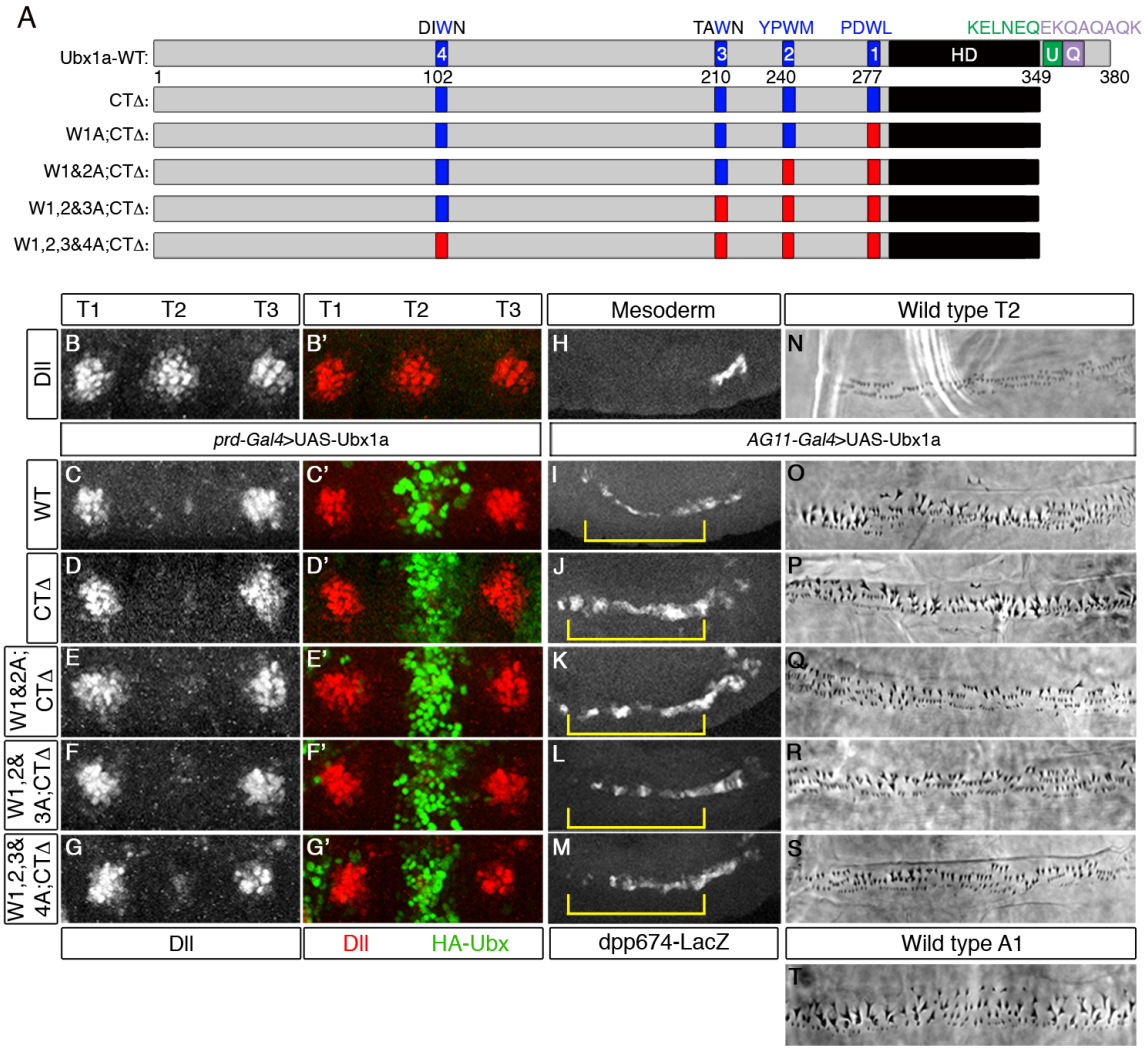


Figure 3.9. Ubx does not require C-terminal or W motifs for in vivo functions. (A) Schematics of wild type and mutant Ubx proteins. Diagrams are approximately to scale. Blue designates W-motifs. Green designates the UbdA motif (U). Purple designates the QA motif (Q). Red indicates residues mutated to alanines (PDWL->AAAA, YPWM->AAAA, TAWN->TAAN and DIWN->DIAN). (B-G) Thoracic region of a wild type embryo (B) or embryos expressing Ubx proteins in T2 via the *prd-Gal4* driver, stained for *Dll* (white or red) and HA-Ubx (green). The protein variant is indicated on the left. (H-M). Visceral mesoderm region of a wild type embryo (H) or embryos expressing Ubx proteins via the *AG11-Gal4* driver, stained for ßgal to monitor *dpp674-lacZ* expression (white). Yellow brackets indicate ectopic *lacZ* activation. (N-T) Phase contrast images depicting ventral T2 cuticle patterns for wild type larvae (N) or animals ectopically expressing wild type Ubx (O) or mutant Ubx variants (P-S). (T) Wild type A1 ventral denticle pattern normally specified by Ubx.

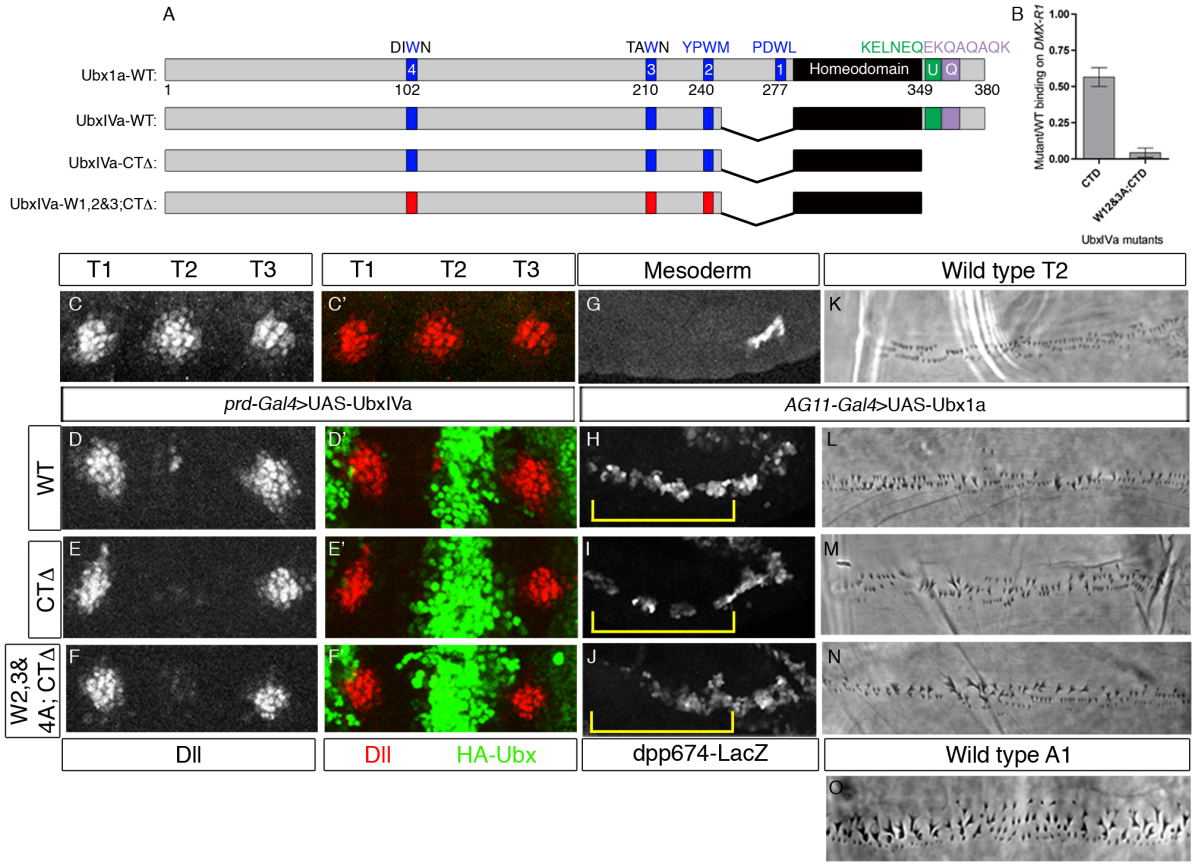


Figure 3.10. Mutation of conserved motifs in UbxIVa does not affect *in vivo* functions despite reducing cooperative complex formation with Exd/Hth *in vitro.* (A) Schematics of wild type and mutant UbxIVa proteins. Diagrams are approximately to scale. Red indicates residues mutated to alanines (YPWM->AAAA, TAWN->TAAN and DIWN->DIAN). (B) Average binding of different UbxIVa mutants to the *DMX-R1* probe. Bar graph represents the mean of n≥3 ratios from independent experiments for each mutant. Error bars represent SEM. (C-F) Thoracic region of a wild type embryo (C) or embryos expressing Ubx proteins in T2 via the *prd-Gal4* driver, stained for *Dll* (white or red) and HA-Ubx (green). The protein variant is indicated on the left. (G-J). Visceral mesoderm region of a wild type embryo (G) or embryos expressing Ubx proteins via the *AG11-Gal4* driver, stained for ßgalactosidase to monitor *dpp674-lacZ* expression (white). Yellow brackets indicate ectopic *lacZ* activation. (K-N) Phase contrast images depicting ventral T2 cuticle patterns for wild type larvae (K) or animals ectopically expressing wild type Ubx (L) or mutant Ubx variants (M and N). (O) Wild type A1 ventral denticle pattern normally specified by Ubx.

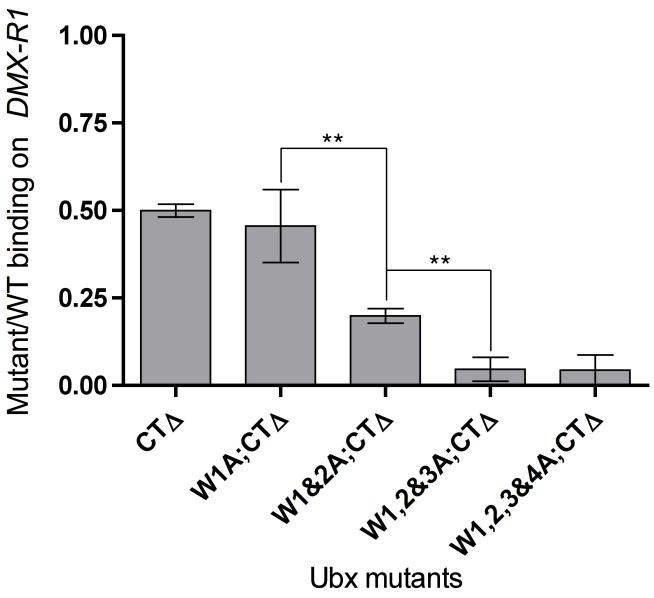


Figure 3.11. Mutation of Ubx W motifs affects cooperative complex formation with Exd-Hth. Average binding of different Ubx mutants to the *DMX-R1* probe. Bar graph represents the mean of n≥3 ratios from independent experiments for each mutant. Error bars represent SEM. T-tests were used to determine if the difference in cooperative binding is significant for a subset of mutants (W1A;CTΔ vs W1&2A;CTΔ p-value=0.003 and W1&2A;CTΔ vs W1,2&3A;CTΔ p-value=0.002).

The Ubx homeodomain and C-terminus are sufficient for several in vivo functions.

Although the sequences C-terminal to the Ubx homeodomain are dispensable for *Dll* repression, sequences N-terminal to the homeodomain are also dispensable: a severely truncated form of Ubx that begins at the homeodomain (Ubx-NTΔ, Figure 3.12A) was a potent repressor of *Dll* (Figure 3.12E). This mutant, which has no W motifs, was also able to activate *dpp674-lacZ* in the visceral mesoderm (Figure 3.12K), but was compromised in its ability to generate a T2 to A1 homeotic transformation (Figure 3.12Q). The activity of this protein is particularly striking, as it only consists of the homeodomain, the UbdA motif, the QA motif, a poly-alanine stretch and eight C-terminal residues that are not conserved (Figure 3.10). When either UbdA or QA was mutated in the NTΔ context, we observed a partial reduction in the ability to repress *Dll* (Figure 3.12F and G). This reduction in repression correlated with a reduction in the ability of these mutants to bind cooperatively with Exd-Hth to *DMX-R1 in vitro* (Figure 3.12B). Additionally, NTΔ;UA and NTΔ;QA induced very weak homeotic transformations

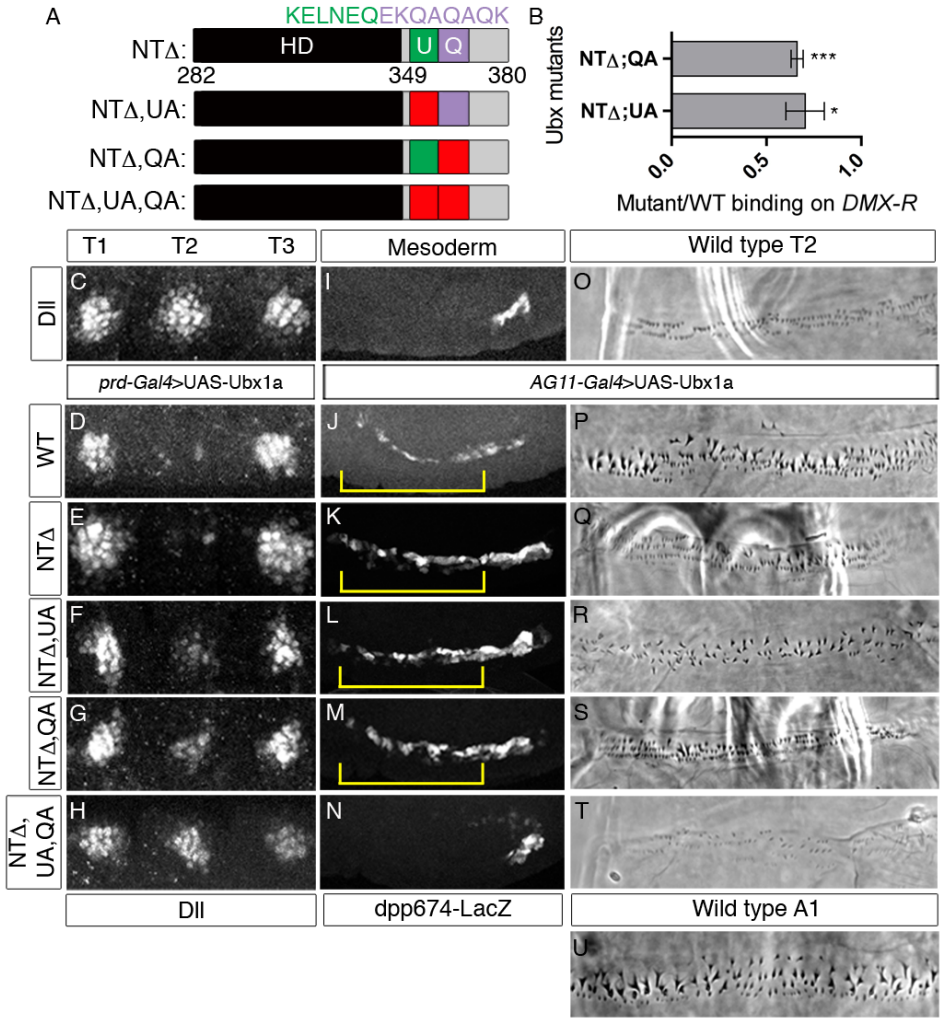


Figure 3.12. C-terminal motifs are sufficient for Ubx function.

(A) Schematics of NTΔ Ubx proteins. Diagrams are approximately to scale. Red indicates residues mutated to alanines or valines (KELNEQ->AALVAV, EKQAQAQK->AAVAVAVA). (B) Average binding of Ubx NTΔ mutants to the *DMX-R* probe. Bar graph represents the mean of n≥3 ratios from independent experiments for each mutant. Error bars represent SEM. T-tests were used to determine if the cooperative binding is significantly different than 1.0 (*** p-value=0.0005 and * p-value=0.043). Due to cleavage of the protein in bacteria, the NTΔ;UA;QA mutant could not be purified and analyzed by EMSA. (C-H) Thoracic region of a wild type embryo (A) or embryos expressing Ubx proteins in T2 via the *prd-Gal4* driver, stained for *Dll* (white or red) and HA-Ubx (green). The protein variant is indicated on the left. (I-N). Visceral mesoderm region of a wild type embryo (I) or embryos expressing Ubx proteins via the *AG11-Gal4* driver, stained for ßgal to monitor *dpp674-lacZ* expression (white). Yellow brackets indicate ectopic *lacZ* activation. (O-T) Phase contrast images depicting ventral T2 cuticle patterns for wild type larvae (O) or animals ectopically expressing wild type Ubx (P) or mutant variants (Q-T). (U) Wild type A1 ventral denticle pattern normally specified by Ubx.

(Figure 3.12R and S) but both were able to activate *dpp674-lacZ* (Figure 3.12L and M). When both UbdA and QA motifs were mutated in the NTΔ context, the protein was completely unable to repress *Dll*, activate *dpp674-lacZ*, and induce Ubx-like homeotic transformations (Figure 3.12H, N and T). Significantly, all of these proteins, even NTΔ;UA;QA, were able to repress *sal* in the wing imaginal disc, demonstrating that they are all still functional transcription factors *in vivo* (Figure 3.13). Together, these data reveal that Ubx is remarkably flexible in its ability to use different motifs to execute its various functions *in vivo*. Moreover, despite many overt similarities with AbdA, Ubx utilizes a unique mechanism for regulating Exd-dependent targets *in vivo*.

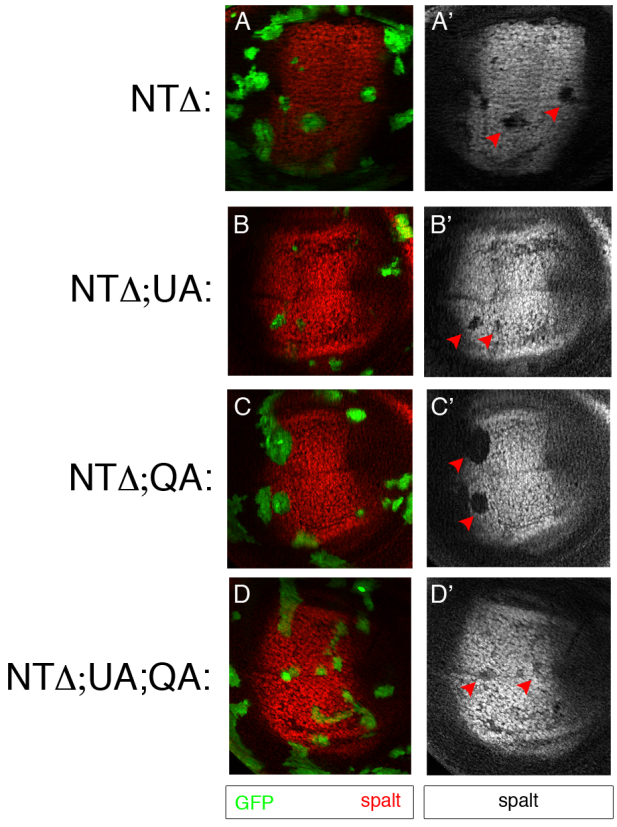
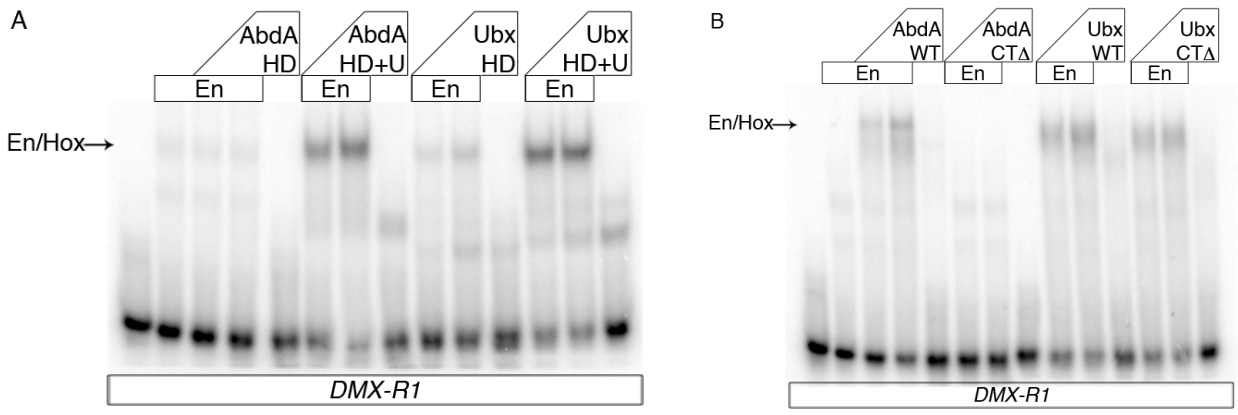


Figure 3.13. N-terminally truncated Ubx proteins repress *sal* in the wing imaginal disc.
Pouch regions of wing imaginal discs containing clones ectopically expressing Ubx or Ubx variants stained for GFP (green, which marks the clones) and Sal (red or white). Representative clones are indicated with red arrowheads.

UbdA mediates cooperative complex formation with En.

Previous studies demonstrated that Exd and Hth are not the only homeodomain proteins that Hox factors can bind DNA cooperatively with: both Ubx and AbdA also bind cooperatively with En to the *Dll DMX-R* element (Gebelein et al., 2004). Using *in vitro* DNA-binding assays, we find that in addition to mediating cooperative complex formation with Exd-Hth (Figures 3.10B and 3.11), the UbdA peptide also contributes to cooperative DNA-binding with En (Figure 3.14). However, there are interesting differences in the way in which these two Hox proteins interact with En. For both Ubx and AbdA, the homeodomain plus UbdA peptide were sufficient to form cooperative complexes with En *in vitro* (Figure 3.14A). However, Ubx, but not AbdA, bound cooperatively with En in the absence of all sequences C-terminal to the homeodomain (Figure 3.14B), suggesting that Ubx has additional sequences in its N-terminus or homeodomain that are sufficient to mediate cooperative binding with En. Together with the Exd binding experiments described above, these results demonstrate that UbdA can mediate complex formation with more than one cofactor. Moreover, they also highlight that the same evolutionarily conserved motif – UbdA – is essential in one context (AbdA) but dispensable in another context (Ubx).

Figure 3.14. The UbdA motif mediates interaction with En.
(A,B) EMSAs of AbdA and Ubx proteins with En on *DMX-R1*. The positions of Hox-En cooperative complexes are indicated with arrows. (A) Truncations include the homeodomain (HD) and/or the UbdA motif (U).



DISCUSSION

Cooperative DNA-binding with cofactors increases the specificity of many transcription factors, including Hox proteins. In the case of Hox and Exd, physical interactions between these factors are critical for complex formation: the conserved YPWM motif, present in nearly all Hox proteins, directly contacts Exd's homeodomain (Mann et al., 2009). However, additional analysis of Hox proteins and their target genes has indicated that the W-Exd interaction is not the only mode of Hox-Exd interaction (Galant et al., 2002; Green et al., 1998; Joshi et al., 2010; Merabet et al., 2003; Merabet et al., 2007; Saadaoui et al., 2011). Indeed, we demonstrate that there is a surprisingly large number of different ways in which Hox proteins functionally interact with cofactors, even when regulating the same target gene.

Increasing specificity using a shared cofactor.

Considering the sequence similarities between the Hox homeodomains and YPWM motifs, it is surprising that the same cofactor can increase DNA-binding and functional specificity for Hox family members. As a counter example, the SOX family of transcription factors uses different cofactors in different developmental contexts to

regulate specific target genes (Kondoh and Kamachi, 2010). Hox proteins are unusual in that they all have the ability to utilize the same cofactor, Exd, when executing many of their *in vivo* functions (Mann and Chan, 1996). One way to increase specificity using a single cofactor may be through changing the mode of interaction. Accordingly, we find that of the three Hox proteins analyzed, each uses a different repertoire of binding mechanisms for regulating Exd-dependent target genes.

Based on our data and previous studies, we suggest that the W motif provides Hox proteins with a basal, shared mechanism for interacting with Exd. Interestingly, other non-Hox proteins, such as En and MyoD, also use tryptophan residues to interact with Exd and Pbx (Knoepfler et al., 1999; Peltenburg and Murre, 1996). In the case of Scr, a single W motif is necessary for Exd-dependent functions. However, the abdominal Hox proteins Ubx and AbdA utilize more complex binding mechanisms. AbdA has an additional W motif, TDWM, as well as sequences in its C-terminus that contribute, in a context dependent manner, to the regulation of Exd-dependent targets. The observation that motifs are differentially required depending on the *in vivo* function suggests that motif utilization may play a role in target site recognition and gene regulation. Previous studies demonstrated that for Scr, interaction of its W motif with the Exd homeodomain helps position paralog specific residues of the linker region in the minor groove of a specific binding site (Joshi et al., 2007). Although additional structural studies are needed to fully address this question, we speculate that the alternative modes of Hox-Exd interaction described here may function in an analogous manner by directly affecting the way Hox proteins interact with DNA. It is also possible that Exd-Hox interactions alter the way in which the homeodomain docks onto DNA. In support of

this idea, previous studies have shown that sequences immediately C-terminal to the homeodomain can play a role in homeodomain structure and specificity (LaRonde-LeBlanc and Wolberger, 2003; Lin and McGinnis, 1992). Consistently, we find that mutating the UbdA motif in Ubx can adversely affect the ability of the homeodomain to bind DNA, even in the absence of Exd (Figure 3.8).

Our results also reveal that despite sharing binding sites and having several similar conserved sequence motifs, Ubx uses a more complex binding mechanism compared to AbdA. Ubx displays impressive flexibility, in that neither sequences N- nor C-terminal to its homeodomain are required for executing some of its functions *in vivo* (Figures 3.9 and 3.12). Ubx's complexity is further enhanced by the possibility of an Exd-independent mechanism based on the ability of some mutants to function *in vivo* despite their inability to cooperatively bind with Exd/Hth *in vitro* (Figures 3.9 and 3.11). However, as our readout for *Dll* expression is protein expressed from the native *Dll* locus (as opposed to a reporter construct), we cannot rule out that Ubx has the ability to bind cooperatively with Exd to other, as yet unidentified, DNA binding sites using sequences still present in the W1,2,3&4A;CTΔ mutant.

In addition to potentially altering Hox-DNA recognition, the fact that some Hox proteins have more and qualitatively distinct cofactor interaction motifs may be relevant to phenotypic suppression (also called posterior dominance), where more posterior Hox proteins post-translationally suppress anterior Hox functions (Bachiller et al., 1994; Gonzalez-Reyes et al., 1990; Mann and Hogness, 1990). Consistent with this notion, the presence of additional posterior-specific cofactor interaction motifs, such as AbdA's

UR motif, endows posterior Hox proteins with the ability to out-compete and phenotypically suppress more anterior Hox proteins (Noro et al., 2011).

Navigating multiple cofactors

Lastly, our data suggest that multiple interaction motifs may help Hox proteins facilitate interactions with other cofactors. In the case of AbdA and Ubx, cooperative complexes on *DMX-R* include both Exd and En. Using *in vitro* DNA-binding assays we found that the UbdA peptide is necessary and sufficient for AbdA to form cooperative complexes with En. Additional structural studies will be necessary to understand the exact mechanism for how a single motif can mediate interaction with multiple cofactors; however, we speculate that having additional Exd-interaction motifs leaves UbdA free to interact with En. Alternatively, UbdA could act as a bridge between the two cofactors, helping to anchor both to the *DMX-R* binding site. Interestingly, UbdA is not required for Ubx to form cooperative complexes with En, again suggesting that the same motif has distinct properties in different Hox proteins. In addition, these results suggest that Ubx has other mechanisms that further enhance its flexibility. Additional homeodomain-containing proteins such as Hth, (and Meis in vertebrates) have been suggested to interact with Hox proteins (reviewed in Moens and Selleri, 2006). Although specific interaction motifs have yet to be identified, our data suggest, at least in the case of AbdA-Hth-Exd on the *rhoA* enhancer, that the motifs examined here are not critical for these interactions. It is curious that although Hth is also a TALE class homeodomain, that the tryptophan-containing motifs are not playing a measurable role in complex formation on this target *in vitro* (Figure 3.5). From the data presented here, it is clear

that the relationship between Hox proteins and their cofactors is not only complex but also critical for how functional specificity is achieved.

MATERIALS AND METHODS

Drosophila strains and genetic manipulations

The GAL4/UAS system (Brand and Perrimon, 1993) was used to ectopically express all UAS-Hox variants in developing embryos. Transgenic UAS lines were all generated using the phi-C31-based integration system (Bischof et al., 2007), except the following AbdA constructs: WT, W1A, W2A, W1&2A, RA and W1&2A;RA, which were generated using standard P-element transformation. All AbdA lines were selected for similar expression levels to each other as detected by western blot. All *phi-C31* based transgenes were cloned into pUAST-attB and epitope tagged. UAS-Scr lines were HA-tagged and inserted into the 51D landing site. UAS-AbdA lines are myc-tagged (Abu-Shaar and Mann, 1998) and those generated using *phi-C31* were also inserted into the 51D landing site. UAS-Ubx lines are all HA-tagged and inserted into the 86Fa landing site. Transgenic lines were confirmed by PCR. Either *AG11-GAL4* on the second chromosome or *prd-GAL4* on the third chromosome was used for ectopic expression as indicated. Scr crosses were grown at 25°C, while AbdA crosses were grown at 28°C (Joshi et al., 2010). Ubx crosses were also grown at 25°C where transgenes were expressed at ≤ 2 fold endogenous Ubx as measured by immunostaining. *fkh250-lacZ* and *dpp674-lacZ* on the third chromosome were used (Capovilla et al., 1994; Ryoo and Mann, 1999). Flip-out clones were generated by crossing *hs-flp; act<y<Gal4, UAS-GFP* to different UAS-Hox lines and heat shocking larvae for 10 minutes at 37°C. Wing discs

were dissected at wandering stage. Cuticle preparations were performed using AG11-GAL4 and prepared as described in Wieschaus and Nusslein-Volhard (Wieschaus, 1986).

Antibodies and immunohistochemistry

Rabbit anti-β-galactosidase (Cappell, 1:5000), mouse anti-Myc (Santa Cruz, 1:200), rat anti-HA (3F10 from Roche, 1:500), mouse anti-Ubx (FP3.38, 1:10) and guinea pig anti-Dll (1:3000) (Estella et al., 2008) were used for staining embryos as previously described (Noro et al., 2006). Rabbit anti-Spalt (1:500) (Xie et al., 2007) was used for imaginal discs immunostaining. Secondary antibodies used were AlexaFluor488 (1:500), AlexaFluor555 (1:1,000) and AlexaFluor647 (1:500) conjugates from Molecular Probes. Embryos and imaginal discs were mounted in Vectashield medium. Z-series were collected on a Leica SP5 confocal microscope. All embryonic images shown are Z-projections of the acquired Z-series.

Protein purification and Electrophoretic Mobility Shift Assays (EMSA)

All proteins were his-tagged and purified from *E.coli* (BL21 or BL21pLysS, Agilent) after 2 hr (Scr and Ubx) or 4 hour (AbdA, Exd-Hth, and Exd-Hth^{HM}) induction with IPTG using either Ni- or Co- chromatography (Gebelein et al., 2002). Constructs were optimized for purification under native conditions: Scr (Joshi et al., 2007), Ubx1a (Ryoo and Mann, 1999), AbdA (Ryoo and Mann, 1999), Exd (Gebelein et al., 2002), Hth (Gebelein et al., 2002), Hth^{HM} (Noro et al., 2006) and En (aa421-552). Truncated proteins not represented in schematics include the following residues: AbdA HD

(aa121-197), AbdA HD+U (aa121-206), Ubx HD(aa282-345), Ubx HD+U (aa282-354), AbdA CTΔ (aa79-197). Mutant variants were made using PCR mutagenesis. Exd-Hth^{HM} and Exd-Hth heterodimers were co-expressed and purified from *E. coli* and used for all the EMSAs (Gebelein et al., 2002). Protein concentrations were determined by the Bradford assay and then confirmed by SDS PAGE and Blue Coomassie analysis (GelCode Blue, Pierce). EMSAs were carried out as previously described (Gebelein et al., 2002). DNA probes include: *fkh250* (Ryoo and Mann, 1999), *DMX-R1* (Gebelein et al., 2004), *DMX-R* (Gebelein et al., 2004), *rhoA* (Li-Kroeger et al., 2008) and *knot* (Hersh et al., 2007). Cooperative DNA binding was calculated as a ratio of the amount of mutant-Exd-Hth complex to the amount of wild type-Exd-Hth complex in the same gel. For each reaction 50ng or 100ng of Exd-Hth^{HM} or Exd-Hth and 150ng of Engrailed were used. Equimolar amounts of Hox proteins were used for each experiment (0.5-1.5pmoles of Scr, 0.5-4.5pmoles of AbdA and 1-15pmoles of Ubx, details upon request). DNA binding was calculated using phosphoimaging as detected by the Typhoon (Amersham) and quantified by ImageQuant (Amersham). Prism (Graphpad software) was used for statistical analyses and to generate graphs.

Addendum 1.

Ubx N-terminal truncations

Ubx has multiple W motifs that coordinate cooperative complex formation with Exd/Hth *in vitro* (Chapter 3). However, none of these motifs are necessary for Ubx function *in vivo* (Chapter 3). To test if any of these motifs are sufficient we constructed a series of N-terminal truncations (Figure A1.1). Unfortunately, many of these mutants either did not express or were mislocalized (Figure A1.1). For example, the construct starting at W3 with all W motifs intact localized to both the nucleus and the cytoplasm (Figure A1.1). However, mutation of W motifs within this truncated context was able to restore full nuclear localization (Figure A1.1). The NLS for all Hox proteins is located in the homeodomain and therefore present in all of these truncations. Based on previous studies that have suggested that the YPWM motif can autoinhibit DNA-binding of Hox proteins (Chan et al., 1996), we believe that the W motifs are masking the NLS via a similar mechanism. Furthermore, these truncated proteins are not detectable via an anti-Ubx antibody where the epitope has been mapped to the homeodomain (data not shown). Therefore, we believe that intra-molecular interactions between the W motifs and the homeodomain are interfering with proper nuclear translocation. In addition to problems with cellular localization, many of the truncations suffer from reduced expression levels. Multiple Gal4 drivers were used with similar results. In attempt to get expression of the homeodomain alone we also constructed a GFP-HD fusion construct. Although this protein is expressed and localized to the nucleus, it is unable to repress Dll (data not shown).

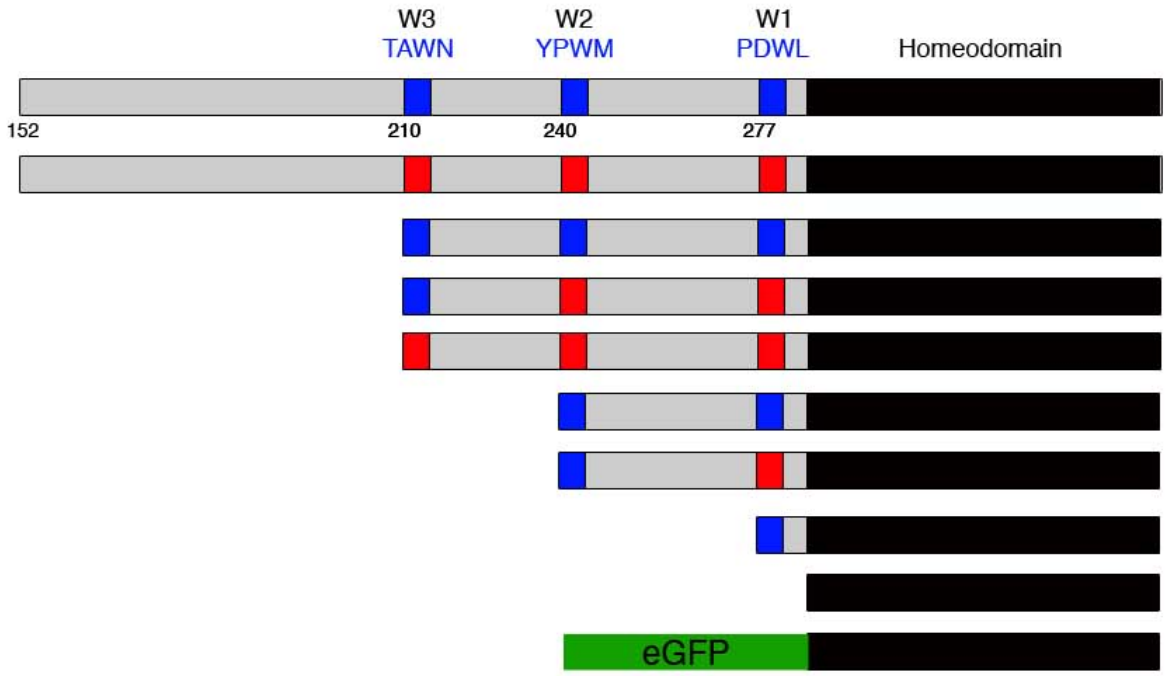


Figure A1.1 Schematics of mutant Ubx proteins.
Blue designates W-motifs. Red indicates residues mutated to alanines (PDWL->AAAA, YPWM->AAAA, and TAWN->TAAN). All constructs are truncated at the end of the homeodomain.

Despite these technical difficulties, several conclusions can be drawn from this data. First, we provide *in vivo* evidence that the W motifs functionally interact with the homeodomain. Furthermore, N-terminal sequences may mitigate these intra-molecular interactions to allow nuclear translocation under wild type conditions. Second, the Ubx homeodomain is not sufficient for *Dll* repression. Lastly, in combination with the C-terminal mutants discussed previously in Chapter 3, these studies suggest that while none of the sequence motifs are necessary they are sufficient for *Dll* repression *in vivo*.

A

Construct	Expression	Localization	Repress *Dll*
+152;CTΔ	~WT	Nuc + Cyto	Yes
+152;W1,2&3A;CTΔ	~WT	Nuc	No
+210;CTΔ	~WT	Nuc + Cyto	Yes
+210;W1&2A;CTΔ	~WT	Nuc	Variable
+210;W1,2&3A;CTΔ	~WT	Nuc	No
+240;CTΔ	<<WT	ND	NA
+240;W1&2A;CTΔ	<<WT	ND	NA
+277;WT;CTΔ	None*	NA	NA
HD	None*	NA	NA
eGFP-HD	ND	Nuc	No

B

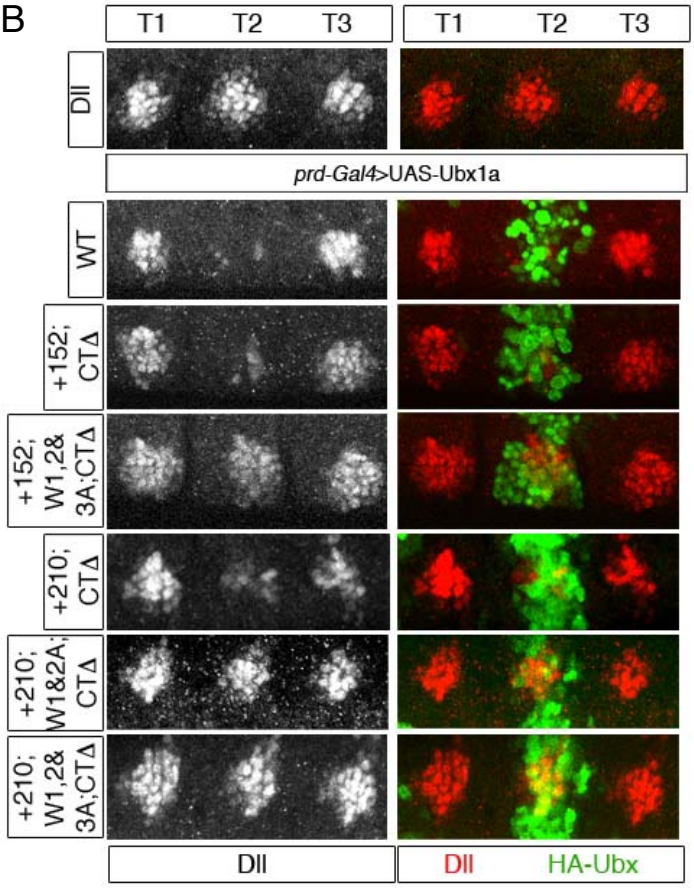


Figure A1.2 Ubx N-terminal truncation mutants do not express or localize properly. A. Table describing results from *in vivo* misexpression experiments. Multiple Gal4 drivers were used. ~WT indicates that the expression levels looked similar to UAS-Ubx^{WT}. However, due to issues with nuclear localization it was difficult to quantify based on immunostaining. <<WT indicates that some expression was seen by eye but was too weak to image. None* means that the levels were below detection via immunostaining. Nuc indicates nuclear localization and Cyto indicates cytoplasmic localization. ND = Not Determined. NA = Not Applicable. B. Thoracic region expressing Ubx proteins in T2 via the *prd-Gal4* driver, stained for *Dll* (white or red) and HA-Ubx (green).

Addendum 2.

Distalless regulation.

Embryonic expression of the transcription factor *Distalless (Dll)* in the leg primordia is necessary for leg development (Cohen, 1990). This early expression of *Dll* is controlled by the 304 enhancer, which contains a Hox-responsive repressor element called *DMX-R* (Gebelein et al., 2004). *DMX-R* has two Hox binding sites flanking Exd and Hth sites (Figure A2.1). Mutation of either Hox site can disrupt reporter expression *in vivo* (Gebelein et al., 2004). However, only the section of *DMX-R* called *DMX-R1*, has the canonical Hox/Exd/Hth binding site arrangement, making it most useful for analyzing the contribution of potential Exd-interaction motifs to cooperative complex formation.

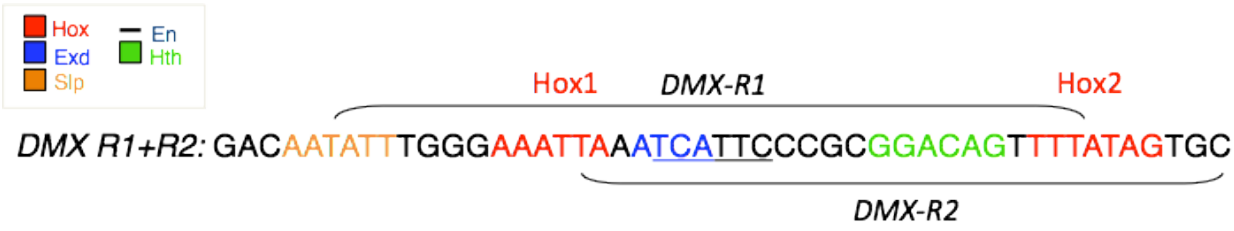


Figure A2.1. Arrangement of binding sites in the *DMX-R* element.
Binding sites are color coordinated and the En binding site is underlined. Brackets indicate sequences included in the *DMX-R1* and *DMX-R2* probes.

This is supported by the correlation between the relative abilities for different AbdA mutants to cooperatively bind *DMX-R1 in vitro* and repress *Dll in vivo* (Chapter 3). However, in the case of *DMX-R2* where the binding site arrangement is Exd/Hth/Hox, mutations in predicted Exd-interaction motifs do not affect cooperative complex formation (Figure A2.2). These data are similar to those on the *rhoA* binding site (Chapter 3), suggesting that the functional interaction may be between AbdA and Hth.

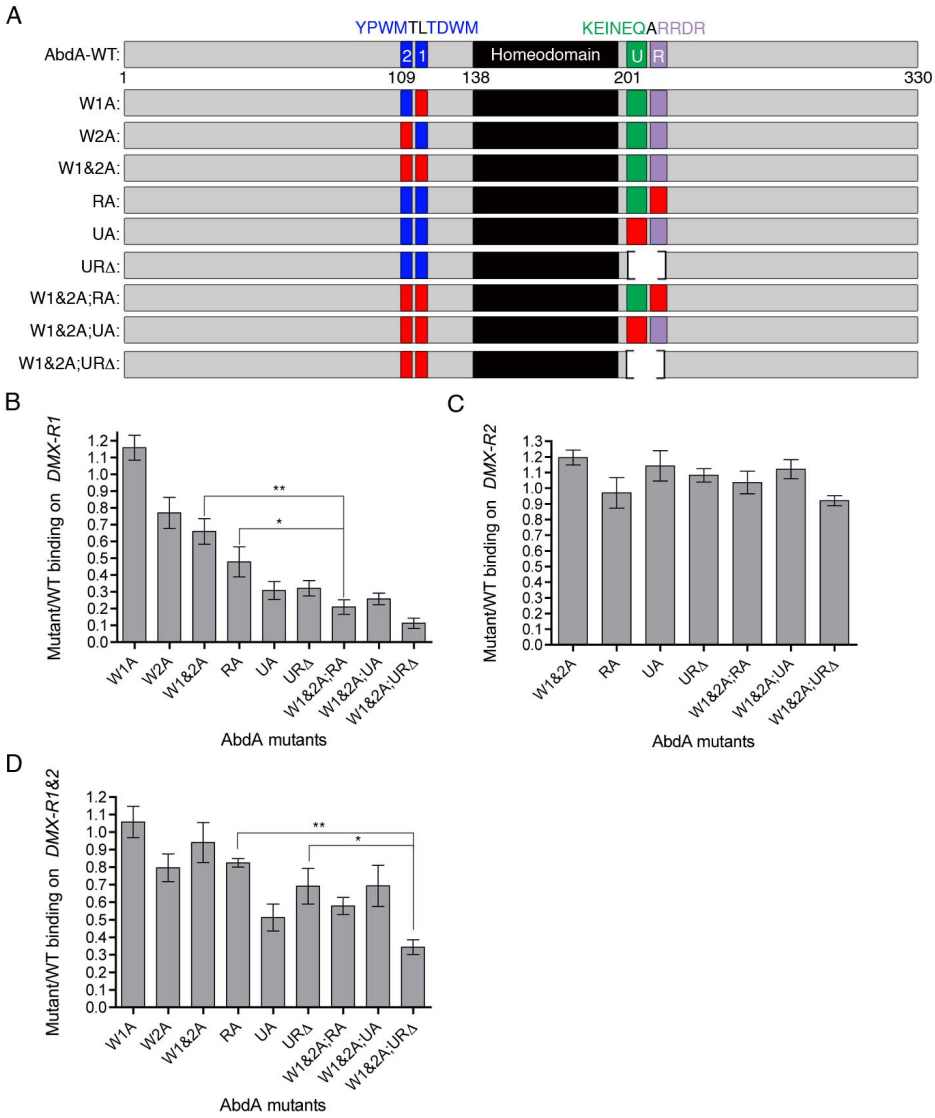


Figure A2.2 AbdA has multiple motifs that mediate cooperative complex formation with Exd/Hth.

(A) Schematics of wild type and mutant AbdA proteins. Diagrams are approximately to scale. Homeodomain (HD) in black. Blue designates W-motifs. Green designates the UbdA motif (U). Purple designates the RRDR motif (R). Red indicates residues mutated to alanines (YPWM->AAAA, TDWM->AAAA, KEINEQ->AAAAAA and RRDR->AAAA). (B) Average binding of different AbdA mutants to the *DMX-R1* probe. T-tests were used to determine if the difference in cooperative binding is significant for a subset of mutants (** p-value=0.004 and * p-value=0.031). (C) Average binding of different AbdA mutants with Hth-Exd to the *DMX-R2* probe. The amount of cooperative complex formation as a ratio of wild type binding did not change significantly across the different mutants (ANOVA analysis: F(6,26)=1.996 p-value=0.1028).(D) Average binding of different AbdA mutants with Hth-Exd to the *DMX-R1&2* probe. T-tests were used to determine if the difference in cooperative binding is significant for a subset of mutants (** p-value=0.008 and * p-value=0.035).

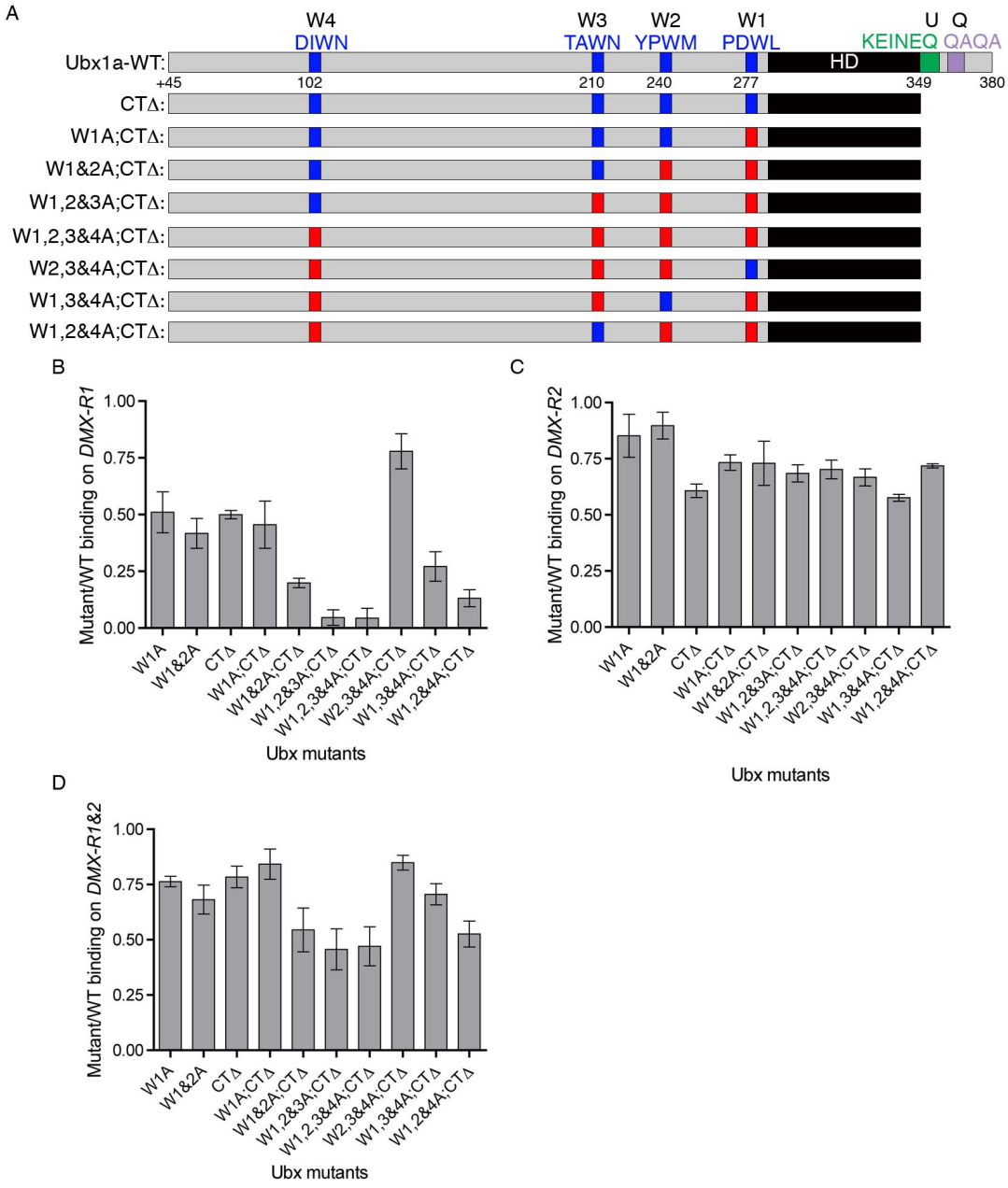


Figure A2.3. Ubx has multiple motifs that contribute to cooperative complex formation with Exd-Hth.

A. Schematics of wild type and mutant Ubx proteins. Diagrams are approximately to scale. Blue designates the W-motif. Green designates the UbdA motif. Purple designates the QA motif. Red indicates residues mutated to alanines (PDWL->AAAA, YPWM->AAAA, TAWN->TAAN and DIWN->DIAN). B. Average binding of different Ubx mutants to the *DMX-R1* probe. C. Average binding of different Ubx mutants to the *DMX-R2* probe. C. Average binding of different Ubx mutants to the *DMX-R1&2* probe. Bar graphs represents the mean of n≥3 ratios from independent experiments for each mutant. Error bars represent SEM.

Binding to the composite *DMX-R1&2* site appears to be an average of the two sites (Figure A2.2).

Analogously, mutations in potential Exd-interaction motifs in Ubx display similar binding profiles to AbdA mutants on *DMX-R1*, *DMX-R2* and *DMX-R1&2 in vitro* (Figure A2.3). Cooperative binding of Ubx mutants with Exd/Hth to *DMX-R1* is drastically affected, while binding to *DMX-R2* is relatively unaffected and binding to the composite site, *DMX-R1&2*, is an average of binding to the two individual sites (Figure A2.3). Since truncation experiments were unable to determine the contribution of individual W motifs (Addendum 3), we made a series of mutants where all the W motifs, except for one, are mutated and the C-terminus is deleted (Figure A2.3). These studies suggest that the W motif closest to the homeodomain is sufficient for cooperative complex formation and the ability to rescue cooperative binding decreases with distance from the homeodomain (Figure A2.3).

In contrast to AbdA, where *in vitro* binding to *DMX-R1* correlates with repression of *Dll in vivo*, all Ubx mutants are able to repress *Dll*, as well as two reporter constructs *304-lacZ* and *DMX-lacZ* (Chapter 3 and Figure A2.4). We propose that Ubx may have an Exd-independent binding mechanism. Future ChIP experiments may address this possibility using *DMX-R* constructs where the Exd/Hth binding sites have ben mutated. However, it is also possible that, unlike AbdA, Ubx can utilize both Hox sites in *DMX-R in vivo*. Experiments designed to test the requirement for each Hox binding site were inconclusive due to differences between repression mechanisms in the anterior and posterior compartments of each segment. *DMX-R* also contains binding sites for the repressor proteins Sloppy paired (Slp) and Engrailed (En) that control *Dll* repression in

the anterior and posterior compartments, respectively (Figure A2.1) (Gebelein et al., 2004). Future ChIP experiments using these mutated constructs may more precisely assess *in vivo* binding.

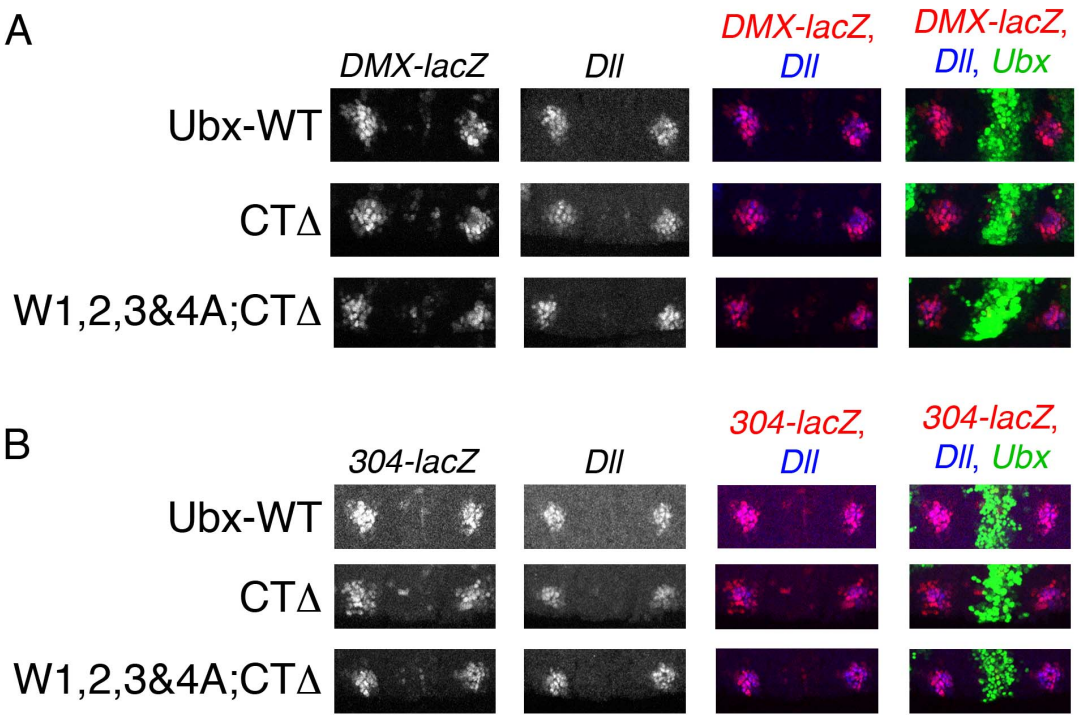


Figure A2.4. Ubx mutants can repress reporter constructs.
A. Ubx mutants are able to repress *DMX-lacZ* and *Dll* expression. *DMX-lacZ* includes the entire *DMX* element as previously described (Gebelein et al., 2004). B. Ubx mutants are able to repress *304-lacZ* and *Dll* expression. *304-lacZ* recapitulates *Dll* expression in early leg primodria as previously described (Vachon et al., 1992). *Prd-Gal4* was used to express *UAS-Ubx* mutants in the second thoracic segment.

Addendum 3.

Extradenticle: Hox is just a ho without it.

The interaction between tryptophan-containing motifs (W motifs) and the TALE motif in the Exd homeodomain is highly conserved (Mann et al., 2009; Moens and Selleri, 2006). The goal of this project was to test the requirement for YPWM-TALE interactions for all Hox proteins *in vivo*. *Exd* is broadly expressed throughout the embryo and it's function is regulated at the level of nuclear translocation by its obligate dimer partner Hth (Rauskolb et al., 1993; Rauskolb and Wieschaus, 1994). Previously, our lab has demonstrated that ubiquitous expression of *exd* using a tubulin driver can rescue *exd* null flies. To take advantage of the *phi-C31* integration system I created an *attP-tubulin-3xFlag-exd* transgene so that each construct could be inserted into the same location to ensure similar expression levels. However, none of the *attP*-based rescue constructs I generated were able to fully rescue *exd* in any of the *attB* landing sites used. Phenotypically, the rescued males were highly uncoordinated and sterile due to non-motile sperm (data not shown). Western blot analyses suggested that the expression levels were insufficient for complete rescue (data not shown).

To test the contribution of the TALE motif to cooperative complex formation *in vitro*, we generated a mutant Exd protein with an internal deletion of the three amino acids in the TALE motif ($Exd^{\Delta LSN}$). Cooperative binding of $Exd^{\Delta LSN}$ with Ubx and AbdA was reduced compared to wild type Exd, and abolished with Scr and Antp (Figure A3.1). These data are consistent with the Hox mutational analyses (Chapters 2 and 3), supporting the hypothesis that Ubx and AbdA have multiple modes of interaction with

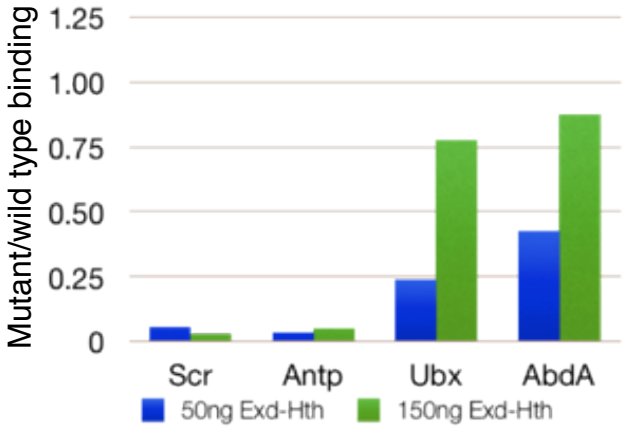


Figure A3.1. Posterior Hox proteins can form complexes with Exd$^{\Delta LSN}$ *in vitro*. Cooperative binding ability of the Exd$^{\Delta LSN}$ mutant. Graph depicts ratios of binding abilities (Exd$^{\Delta LSN}$-Hox/Exd^{wt}-Hox complexes) over several independent experiments. EMSAs were conducted as described previously using the consensus probe *fkh250^{CON}* (Gebelein et al., 2002; Ryoo and Mann, 1999).

Exd, while Scr and Antp rely on a single W motif. One caveat to these experiments is that the DNA binding of the Exd$^{\Delta LSN}$-Hth dimer is less than 1/4 of wild type dimers (Figure A3.2 green bars). In an attempt to create a TALE mutant of Exd that would maintain close to wild type levels of DNA binding ability but interfere with W motif interactions, additional constructs were designed and tested *in vitro*. Unfortunately, all of the mutants that demonstrated improved DNA binding (compared to the Exd$^{\Delta LSN}$ mutant) showed a similar increase in complex formation (Figure A3.2). After close inspection of the Exd homeodomain structure it appears that the conformation of the TALE motif is critical for positioning an adjacent 'PYP' motif to directly contact the DNA phosphate backbone (Figure A3.3). Although the 'PYP' motif does not make base-specific contacts, this interaction may help anchor the Exd homeodomain to the DNA.

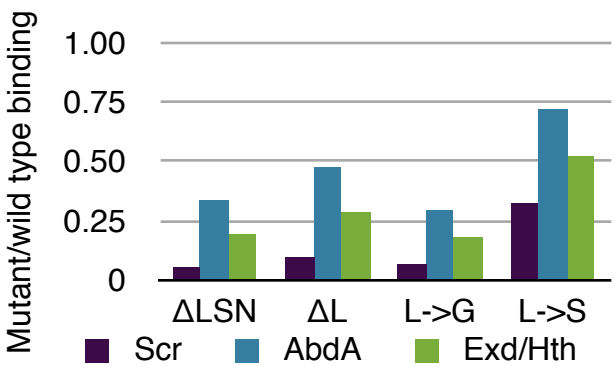


Figure A3.2. Mutations in the TALE motif affect DNA binding. Graph depicts ratios of binding abilities for different Exd TALE mutants with different Hox proteins or as a dimer with Hth. EMSAs were conducted as described previously using the consensus probe *fkh250^{CON}* (Gebelein et al., 2002; Ryoo and Mann, 1999).

Consistently, when the region between the second and the third helices of the Exd homeodomain was replaced with the corresponding region of a consensus homeodomain structure lacking the 'PYP' motif, DNA binding was completely lost (data not shown). From these experiments and observations we concluded that due to structural constraints the ideal Exd mutant (one that interferes with W motif interactions while maintaining DNA binding) may not be possible.

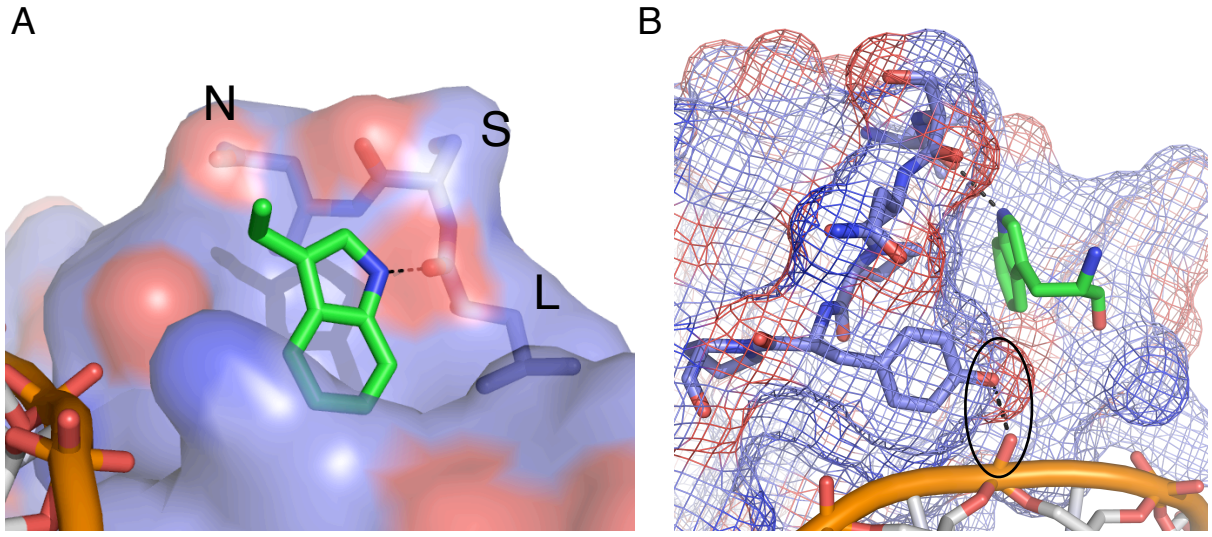


Figure A3.3. Structure of the Exd homeodomain bound to DNA.
A. Hydrophobic pocket created by the TALE motif (LSN residues indicated). B. TALE-PYP (LSNPYP) represented as sticks and the mesh represents the surface structure of the Exd homeodomain. Oval highlights the polar contact of Tyr25 with the phosphate backbone of DNA. Tryptophan of Scr shown in green. Polar contacts are indicated by dashed black lines. PDB from (Joshi et al., 2007).

The identification of non W motifs that mediate cooperative complex formation with Exd suggested that Exd might harbor additional Hox interaction motifs outside of the TALE motif in the homeodomain (Chan and Mann, 1993; Lelli et al., 2011; Merabet et al., 2011; Merabet et al., 2007; Noro et al., 2011; Saadaoui et al., 2011). We sought to identify additional sequences in Exd that contribute to Hox-Exd cooperative DNA binding. DNA binding experiments using AbdA and different N-terminal truncations of

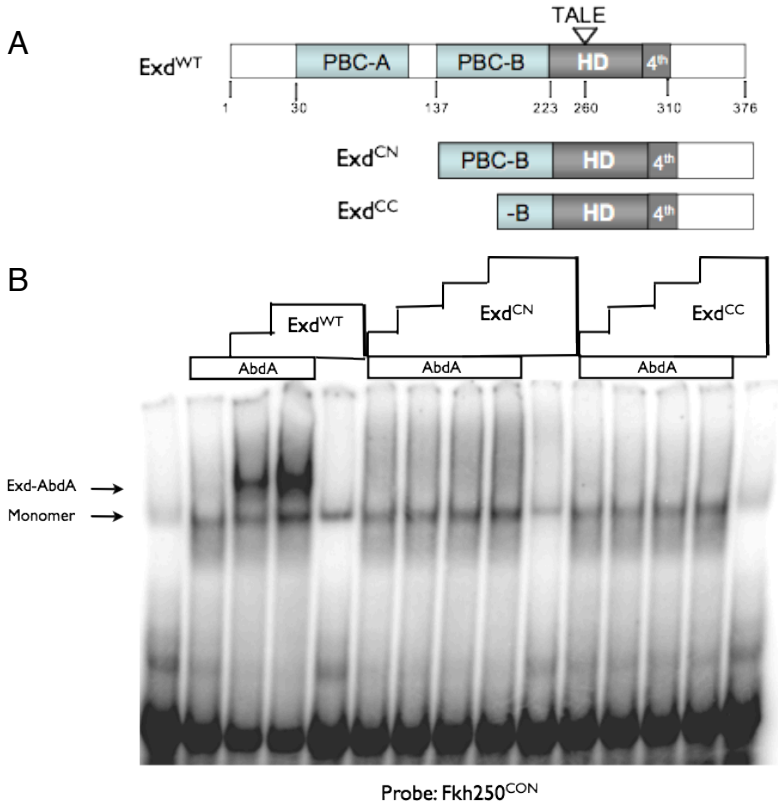


Figure A3.4. Exd's PBC-A is required for cooperative DNA binding with AbdA. A. Schematics of Exd proteins used. 4th indicates an additional helix that contributes to homeodomain (HD) function. B. EMSAs were conducted as described previously using the consensus probe *fkh250*^{CON} (Gebelein et al., 2002; Ryoo and Mann, 1999).

Exd suggested that sequences in the PBC-A domain were required for cooperative complex formation (Figure A3.4). Based on these results we conducted an alanine-scan of the PBC-A domain assaying interaction with the HM domain of Hth and cooperative complex formation with Scr, Ubx and AbdA (Figure A3.5). We also tested the effect of deleting a predicted coiled-coil domain (CC) in the PBC-B domain. The heat map in Figure A3.4 summarizes the results of these experiments, where green indicates cooperative complex formation similar to wild type and red indicates a loss of cooperative binding. All mutants were able to interact with the HM domain of Hth as tested by co-purification from *E. coli*. All of the mutants that displayed a reduction in or loss of cooperative complex formation (mutants 8, 9, 14 and CC) correlated with a reduction in or loss of DNA-binding as a heterodimer with Hth (data not shown). Unfortunately, due to the effect these mutations have on DNA-binding it is not possible

to discern effects on cooperative complex formation. To test Exd-Hox interactions off the DNA, we attempted a series of pulldown experiments using both GST (Glutathione S-transferase) and MBP (maltose binding protein). However, these experiments were uninformative due to a general loss of interaction specificity (data not shown), suggesting that DNA-binding contributes to Hox-Exd interactions.

A N-SISEQ^QARKHTLNCHRMKPALFSVLCEIKEKTVLSIRNTQEEEPPDPQLMRLDNM-C

B

	3	5	6	7	8	9	10	11	12	13	14	CC
Scr												
Ubx												
AbdA												

Figure A3.5. Mutations in the PBC domains of Exd affect DNA binding.
A. Sequences in the PBC-A domain that were mutated to alanines. Each sequential section of five amino acids is colored coordinated with a number in the table. B. Heat map displaying experimental results. Dark green indicates binding levels similar to wild type. Light green indicates a reduction in binding compared to wild type. Red indicates a loss of binding. CC is an internal deletion of the predicted coiled-coil in PBC-B.

Chapter 4

Discussion and future directions

Hox proteins have fascinated researchers for decades. Mutations in *Drosophila* Hox genes produce dramatic phenotypes such as the famous antenna to leg transformations (Gehring, 1987). At a molecular level, they are equally intriguing, displaying distinct functional specificities despite the potential for promiscuous DNA-binding. The classic example of this "specificity paradox," Hox genes provide a valuable system for studying transcription factor function and gene regulation. The goal of my work was to understand how Hox proteins use cofactors, such as the PBC proteins Extradenticle and Homothorax, to increase their functional specificity. The mutational analyses described here have provided us with an expanded view of how Hox proteins functionally interact with cofactors. However, as with most research projects, this work poses more questions than it was originally designed to address. How do different motifs contribute to specificity? How do complexes using different modes of interaction differ structurally? What is the evolutionary relevance of having multiple motifs? How do interactions with cofactors other than Exd contribute to *in vivo* binding of Hox proteins?

How could different modes of interaction contribute to specificity?

Recent studies from our lab demonstrate that cooperative complex formation with Exd reveals latent specificities of Hox proteins (Slattery et al., 2011). One way this is accomplished depends on interaction with Exd to position residues between the YPWM

and the homeodomain, called the specificity module, to recognize the sequence-dependent DNA structure in the minor groove of specific targets (Joshi et al., 2007; Rohs et al., 2010). Therefore, in the case of Scr and Dfd where a single W motif is required for Exd-dependent functions, specificity is likely dictated by the homeodomain and the specificity module (Joshi et al., 2010). However, other mechanisms may exist for for Hox proteins that have additional modes of cofactor interaction. For example, based on a series of point mutations in the Ubx UbdA motif (Chapter 3), we postulate that the UbdA motif contributes to DNA-binding by either modulating homeodomain function or directly participating in DNA recognition. Furthermore, the RRDR motif of AbdA and the QA motif of Ubx are both positioned close to the UbdA motif and may function in a similar manner, perhaps to modulate the function of the UbdA motif. It has been suggested that motifs may functionally interact to control specificity (Merabet et al., 2011; Saadaoui et al., 2011). Our data is not inconsistent with this idea. However, until we have a better structural understanding of how these motifs interact with cofactors and DNA it will be difficult to say how these motifs contribute to specificity. Therefore, future structural studies should use larger and perhaps mutant proteins on different target sites to address these issues.

Additionally, SELEX-seq experiments are under way to determine how mutating different motifs affects the specificity of Ubx and AbdA. Based on the data described here, we would predict that different subsets of binding sites would be recognized depending on the mode of interaction. However, the current SELEX-seq data available uses a homeodomain-less isoform of Hth called HM. Although the HM isoform is sufficient for many Hox-related functions, including activation of *fkh250-lacZ* (Noro et al.,

2006), future experiments using full-length Hth will provide a more complete list of possible Hox binding sites by including sites that require Hth DNA-binding, such as the *DMX-R* and different binding site arrangements such as Hox/Hth/Exd. These sites would provide helpful tools for characterizing Hox-Hth interactions. Including the Hth homeodomain may also increase the flexibility of the complex, allowing for all modes of interaction to contribute to specificity. This may be particularly important for Hox proteins such as Ubx and AbdA that require Hth binding for many of the currently identified target sites (Gebelein et al., 2004; Li-Kroeger et al., 2008). In vertebrates, the Hth homologs Meis and Prep also contribute to Hox functions (Moens and Selleri, 2006). Therefore, understanding how Hth and it's homologs contribute to binding site selection would provide further insight into how complex flexibility contributes to specificity. Lastly, comparing the Exd/HM and Exd/Hth SELEX-seq data sets would determine the contribution of Hth to the binding specificities of different Hox proteins. For example, Hox proteins with only a single mode of interaction, such as Scr and Dfd, should have a smaller, more restricted set of binding sites; whereas Hox proteins with multiple modes of interaction should have a larger, more diverse array of possible binding sites. Assaying different Ubx and AbdA mutants in this context may help to parse out which configurations contribute to binding at different sites.

Although these *in vitro* experiments would provide valuable information regarding binding site composition, locating relevant binding sites *in vivo* is challenging with such short sequences. The generation of different genomic rescue constructs would provide powerful tools to address specificity *in vivo.* First, these constructs could be used to conduct more comprehensive phenotypic analyses by avoiding the lethality associated

with ectopic misexpression of Hox genes using the Gal4-UAS system. Moreover, barring any dominant negative effects, combining genomic rescue constructs with the flp-FRT system would allow for tissue specific rescue of individual Hox proteins at endogenous levels. Second, if constructed with an efficient epitope tag these constructs could also be used for ChIP-seq or ChIP-exo analysis to identify target sites (Rhee and Pugh, 2011). Furthermore, pairing these experiments with tissue specific expression of a nuclear tag could provide additional refinement using nuclear immunoprecipitation to first isolate nuclei from specific cells.

Comparing Ubx and AbdA

The posterior Hox proteins Ubx and AbdA are similar in many aspects. First, both are abdominal Hox proteins that have some overlapping targets such as *Distalless* and *dpp* (Capovilla et al., 1994; Vachon et al., 1992). Second, their homeodomains differ by only three amino acids. SELEX-seq results demonstrate that they recognize similar binding sites (Slattery et al., 2011). Third, both are able to repress wing genes such as *spalt* (Galant et al., 2002). Lastly, both have multiple motifs that mediate cooperative complex formation with Exd *in vitro* (Chapters 2 and 3). However, despite many of these similarities AbdA and Ubx behave very differently in the mutational analyses described here. For example, while the overall *in vitro* binding profiles of mutant proteins are similar (Addendum 2), corresponding mutants display dramatically different functionalities *in vivo* (Chapter 3). One very simple explanation for these results could be that the truncated Ubx proteins used for *in vitro* analyses do not adequately represent the binding capabilities of the wild type Ubx protein *in vivo*. Another

explanation could be that Ubx has an Exd-independent binding mechanism or that other binding mechanisms not affected by the mutations function redundantly. Future ChIP experiments analyzing Ubx binding at sites where the Exd/Hth sites have been mutated would determine if Ubx can bind without Exd/Hth. However, determining if other binding mechanisms are contributing to Ubx function will be more challenging due to differences between repression complexes in the anterior vs posterior compartments. Mutational analyses of the *DMX-R* element demonstrate that repression in the posterior compartment is highly sensitive to mutations, whereas the anterior compartment is more resilient (Gebelein et al., 2004). Furthermore, based on the observations that En can cooperatively interact with Ubx and AbdA, we speculate that repression in the posterior compartment is mediated by a highly cooperative protein complex containing: Hox, En, Exd and Hth and that perturbation of this complex in any way can result in derepression. In contrast, no cooperative binding is observed between Slp and Hox proteins, suggesting that these complexes may be more flexible (Gebelein and Mann, 2007). Furthermore, Ubx is only expressed in the anterior compartments of the abdomen suggesting that evolutionary pressures on Ubx and AbdA may differ with respect to *Dll* regulation. Therefore, it is possible that because AbdA is responsible for repressing *Dll* in the posterior compartments it most effectively utilizes the first Hox site that is adjacent to the En binding site. This hypothesis is supported by the correlation between *in vitro* binding on *DMX-R1* and *Dll* repression *in vivo* (Chapter 3). Similarly, since Ubx is only expressed in anterior compartments and coordinates *Dll* repression with Slp, Ubx may be able to more efficiently utilize both Hox binding sites. This hypothesis would provide one explanation for why, despite similar binding abilities on *DMX-R2 in vitro* (Addendum

2), only Ubx mutants are able to repress *Dll in vivo*. Future studies aimed at understanding the difference between the binding mechanisms in the anterior vs posterior compartments would help to discern how Ubx vs AbdA and En vs Slp control *Dll* regulation.

Despite these discrepancies regarding *Dll* repression, AbdA is more posterior to Ubx and, according to phenotypic suppression, will dominate over Ubx. Based on the data presented, the quantity and quality of AbdA's Exd-interaction motifs should outcompete Ubx for binding and regulation of shared targets. Since both proteins repress *Dll* it is not possible to test this hypothesis in the context of *DMX-R*; however, Ubx activates expression of *fkh250^{CON}-lacZ*. Future *in vivo* competition experiments would address if the same principles that applied to Scr and AbdA apply to two Hox proteins with multiple motifs. However, one caveat to these experiments, and those presented in Chapter 2, is that we cannot control for the possibility that repressive mechanisms dominate over activation. Ideally, a reciprocal system where the more anterior protein represses and the posterior protein activates target gene expression via the same binding site would address this caveat. Unfortunately, no such enhancer exists to date. Hopefully, future studies will identify a site suitable for these studies.

Vertebrate Hox-cofactor interactions

Unlike *Drosophila,* which has a single copy of each Hox protein, vertebrates have multiple paralogous copies of the Hox clusters with a total of 39 Hox genes. Additional copies of the Hox cofactors including: four Pbx genes, the ortholog of Exd, four Meis and two Prep genes, the Hth orthologs (Moens and Selleri, 2006), further increase the

potential complexity of the vertebrate Hox-cofactor code. However, due to a general lack of direct targets and complications from genetic redundancy, most studies analyzing Hox-cofactor interactions have focused on the genetic requirement for tryptophan containing motifs. All vertebrate Hox proteins, except some of the AbdB orthologs in paralog groups 11-13, have W motifs that facilitate interactions with Pbx proteins (Moens and Selleri, 2006). Similar to Scr and Dfd, some vertebrate Hox proteins require a W motif for some functions (Delval et al., 2011; Knoepfler et al., 2001; Remacle et al., 2004; Schnabel et al., 2000; Vitobello et al., 2011; Yaron et al., 2001). For example, targeted mutagenesis of the murine Hoxa1 W motif parallel the *Hoxa1* loss-of-function phenotype during hindbrain-patterning (Remacle et al., 2004). However, other functions such as neural crest cell migration and cranial nerve development are not as severely affected as the *Hoxa1* loss-of-function animals, suggesting that the W motif is required in a context dependent manner (Remacle et al., 2004). Other studies have also demonstrated differential requirements for W motifs (Fischbach et al., 2005; Knoepfler et al., 2001; Shen et al., 2004). In the case of *Hoxb6*, overexpression of either wild type or W motif mutant proteins in bone marrow cells increases cell proliferation and self renewal as well as formation of immortalized myeloid cell lines (Fischbach et al., 2005). However, co-expression of *Meis1* synergizes with *Hoxb6* overexpression during acute myloid leukemia progression (Fischbach et al., 2005). Therefore, the functional interaction may be between Hoxb6 and Meis1 in a similar manner to AbdA and Hth on the *rhoA* binding site (Li-Kroger et al., 2006). Identification of direct targets and binding sites will address whether cooperative complex formation between Hobx6 and Meis can explain the W motif-independent function of Hoxb6. It has also been

shown using yeast two-hybrid assays that some Hox proteins can interact with Meis proteins, further supporting the possibility that Hox-Meis interactions could contribute to some Hox functions (Williams et al., 2005). In the case of Hoxb8, mutation of the W motif causes dominant homeotic transformations similar to loss-of-function phenotypes for other Hox proteins (Medina-Martinez and Ramirez-Solis, 2003). One explanation could be that removal of the autoinhibitory effect of the W motif may allow Hoxb8 mutant proteins to outcompete other Hox proteins for DNA-binding (Medina-Martinez and Ramirez-Solis, 2003). The ability for mutant Hox proteins to display hyperactive functions when the W motif is mutated has also been observed for the *Drosophila* Hox protein Labial (Chan et al., 1996). Identification of direct binding sites would allow for ChIP analyses to test this hypothesis. Additionally, data from hematopoietic stem cells suggest that regions outside of the W motif and the homeodomain contribute to Hox4-induced self-renewal (Iacovino et al., 2009). These studies and those described in this thesis argue that multiple motifs often contribute to Hox/cofactor cooperative binding and function. Therefore, the existence of non W-motifs must be addressed before declaring that a Hox readout is or is not cofactor-independent. Further studies are also needed to determine the contribution of Hox-Meis and Hox-Prep interactions to Hox specificity.

A recent study using different ChIP-based assays suggests that Pbx1 may function as a pioneer factor for Estrogen Receptor alpha (ERα) recruitment to known binding sites in breast cancer cells (Magnani et al., 2011). Depletion of Pbx1 using short-interfering RNAs decreased ERα binding at sites predicted to have Pbx1 pre-loaded on the DNA (Magnani et al., 2011). Binding of ERα at sites predicted to be Pbx1-

independent were largely unaffected (Magnani et al., 2011). Although it is unknown if the ability for Pbx to function as a pioneer factor affects Hox-related functions, it is possible that alterations in DNA accessibility may be another way cofactors can affect Hox target specificity. To date it is unknown if the *Drosophila* homolog Exd is also capable of acting as pioneer factor. Future studies will hopefully address the molecular mechanisms allowing Pbx and/or Exd to function as both pioneer factors and cooperative cofactors and how these two functions can work together to regulate gene expression.

Hox specificity in the era of genome-wide studies.

Recent genome-wide studies demonstrating widespread binding of site-specific transcription factors highlight the role of DNA accessibility and question the importance of DNA-binding specificity in gene regulation (Biggin, 2011; Li et al., 2008; MacArthur et al., 2009; MacQuarrie et al., 2011). Some members of the field suggest that instead of discrete binding events at specific target loci, transcription factors bind to the majority of genes and that the level of occupancy correlates with regulation (Biggin, 2011). While DNA accessibility likely significantly contributes to transcription factor binding, several lines of evidence argue against the 'Quantitative Continua' hypothesis. First, models generated using genome-wide binding data suggest that within regions of open chromatin transcription factor binding can be predicted based on sequence specificity (Kaplan et al., 2011). Second, many of these studies use whole embryo ChIP-chip data, averaging binding across many dynamic tissue types. Using tissue-specific ChIP-exo may refine the number the of binding events (Rhee and Pugh, 2011). Lastly, as described here, small mutations outside of the DNA binding domain can have significant

effects on *in vivo* regulation. If specificity does not contribute to target gene regulation, than activity regulation should not be affected as long as the innate DNA-binding ability of transcription factors is preserved. As demonstrated by clonal repression of the wing gene *spalt* (Figures 2.5, 3.2, 3.6, 3.13) all of the mutants generated are able to bind and regulate target genes. Furthermore, mutations in different sequence motifs, outside of the homeodomain, affect gene regulation in a site specific manner (Chapter 3). These data all suggest that DNA-binding specificity contributes to transcription factor target gene regulation and argues against a widespread quantitative model.

Although the work described here focuses on only a subset of *Drosophila* Hox proteins, it begins to address mechanisms by which a relatively small number of transcription factors can regulate thousands of genes in response to a myriad of intra- and intercellular signals. Hopefully, future research will build upon this work to further integrate different mechanisms of transcription factor specificity into the complex process of gene regulation.

References

Abu-Shaar, M., and Mann, R.S. (1998). Generation of multiple antagonistic domains along the proximodistal axis during Drosophila leg development. Development *125*, 3821-3830.

Abu-Shaar, M., Ryoo, H.D., and Mann, R.S. (1999). Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. Genes & development *13*, 935-945.

Agarwal, P., Verzi, M.P., Nguyen, T., Hu, J., Ehlers, M.L., McCulley, D.J., Xu, S.M., Dodou, E., Anderson, J.P., Wei, M.L.*, et al.* (2011). The MADS box transcription factor MEF2C regulates melanocyte development and is a direct transcriptional target and partner of SOX10. Development *138*, 2555-2565.

Agelopoulos, M., McKay, D.J., and Mann, R.S. (2012). Developmental regulation of chromatin conformation by Hox proteins in *Drosophila*. Cell Reports, in press.

Arnosti, D.N., and Kulkarni, M.M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? Journal of cellular biochemistry *94*, 890-898.

Bachiller, D., Macias, A., Duboule, D., and Morata, G. (1994). Conservation of a functional hierarchy between mammalian and insect Hox/HOM genes. The EMBO journal *13*, 1930-1941.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X.*, et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720-1723.

Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. Trends in genetics : TIG *26*, 476-483.

Bai, L., Ondracka, A., and Cross, F.R. (2011). Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter. Molecular cell *42*, 465-476.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research *21*, 381-395.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends in genetics : TIG *27*, 454-464.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214-226.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays : news and reviews in molecular, cellular and developmental biology *34*, 135-141.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barth, T.K., and Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. Trends in biochemical sciences *35*, 618-626.

Barthel, R., Tsytsykova, A.V., Barczak, A.K., Tsai, E.Y., Dascher, C.C., Brenner, M.B., and Goldfeld, A.E. (2003). Regulation of tumor necrosis factor alpha gene expression by mycobacteria involves the assembly of a unique enhanceosome dependent on the coactivator proteins CBP/p300. Molecular and cellular biology *23*, 526-533.

Benayoun, B.A., and Veitia, R.A. (2009). A post-translational modification code for transcription factors: sorting through a sea of signals. Trends in cell biology *19*, 189-197.

Bender, W. (2008). MicroRNAs in the Drosophila bithorax complex. Genes & development *22*, 14-19.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Bernard, P., and Harley, V.R. (2010). Acquisition of SOX transcription factor specificity through protein-protein interaction, modulation of Wnt signalling and post-translational modification. The international journal of biochemistry & cell biology *42*, 400-410.

Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.H., Trump, S., Lightman, S.L.*, et al.* (2011). Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Molecular cell *43*, 145-155.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Developmental cell *21*, 611-626.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nature genetics *44*, 148-156.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Busser, B.W., Shokri, L., Jaeger, S.A., Gisselbrecht, S.S., Singhania, A., Berger, M.F., Zhou, B., Bulyk, M.L., and Michelson, A.M. (2012). Molecular mechanism underlying the regulatory specificity of a Drosophila homeodomain protein that specifies myoblast identity. Development *139*, 1164-1174.

Capovilla, M., Brandt, M., and Botas, J. (1994). Direct regulation of decapentaplegic by Ultrabithorax and its role in Drosophila midgut morphogenesis. Cell *76*, 461-475.

Carroll, S.B., Laymon, R.A., McCutcheon, M.A., Riley, P.D., and Scott, M.P. (1986). The localization and regulation of Antennapedia protein expression in Drosophila embryos. Cell *47*, 113-122.

Cavalli, G. (2007). Chromosome kissing. Current opinion in genetics & development *17*, 443-450.

Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R.S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell *78*, 603-615.

Chan, S.K., and Mann, R.S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. Genes & development *7*, 796-811.

Chan, S.K., Popperl, H., Krumlauf, R., and Mann, R.S. (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. The EMBO journal *15*, 2476-2487.

Charlot, C., Dubois-Pot, H., Serchov, T., Tourrette, Y., and Wasylyk, B. (2010). A review of post-translational modifications and subcellular localization of Ets transcription factors: possible connection with cancer and involvement in the hypoxic response. Methods Mol Biol *647*, 3-30.

Chopra, V.S., and Mishra, R.K. (2006). "Mir"acles in hox gene regulation. BioEssays : news and reviews in molecular, cellular and developmental biology *28*, 445-448.

Cohen, S.M. (1990). Specification of limb development in the Drosophila embryo by positional cues from segmentation genes. Nature *343*, 173-177.

Courey, A.J. (2001). Cooperativity in transcriptional control. Current biology : CB *11*, R250-252.

Dai, Z., and Dai, X. (2012). Nuclear colocalization of transcription factor target genes strengthens coregulation in yeast. Nucleic acids research *40*, 27-36.

Daitoku, H., Sakamaki, J., and Fukamizu, A. (2011). Regulation of FoxO transcription factors by acetylation and protein-protein interactions. Biochimica et biophysica acta *1813*, 1954-1960.

Dasen, J.S., and Jessell, T.M. (2009). Hox networks and the origins of motor neuron diversity. Curr Top Dev Biol *88*, 169-200.

Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911-920.

Davidson, E.H., and Levine, M.S. (2008). Properties of developmental gene regulatory networks. Proceedings of the National Academy of Sciences of the United States of America *105*, 20063-20066.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes & development *26*, 11-24.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Delval, S., Taminiau, A., Lamy, J., Lallemand, C., Gilles, C., Noel, A., and Rezsohazy, R. (2011). The Pbx interaction motif of Hoxa1 is essential for its oncogenic activity. PloS one 6, e25247.

Dostie, J., and Bickmore, W.A. (2012). Chromosome organization in the nucleus - charting new territory across the Hi-Cs. Current opinion in genetics & development.

Duboule, D. (1991). Patterning in the vertebrate limb. Current opinion in genetics & development *1*, 211-216.

Duboule, D., and Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. Trends in genetics : TIG *10*, 358-364.

Duncan, I. (1987). The bithorax complex. Annual review of genetics *21*, 285-319.

Duncan, I.W. (2002). Transvection effects in Drosophila. Annual review of genetics *36*, 521-556.

Dunipace, L., Ozdemir, A., and Stathopoulos, A. (2011). Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. Development *138*, 4075-4084.

Enriquez, J., Boukhatmi, H., Dubois, L., Philippakis, A.A., Bulyk, M.L., Michelson, A.M., Crozatier, M., and Vincent, A. (2010). Multi-step control of muscle diversity by Hox proteins in the Drosophila embryo. Development *137*, 457-466.

Ernst, P., Mabon, M., Davidson, A.J., Zon, L.I., and Korsmeyer, S.J. (2004). An Mll-dependent Hox program drives hematopoietic progenitor expansion. Current biology : CB *14*, 2063-2069.

Eskiw, C.H., and Fraser, P. (2011). Ultrastructural study of transcription factories in mouse erythroblasts. Journal of cell science *124*, 3676-3683.

Estella, C., McKay, D.J., and Mann, R.S. (2008). Molecular integration of wingless, decapentaplegic, and autoregulatory inputs into Distalless during Drosophila leg development. Developmental cell *14*, 86-96.

Ethier, S.D., Miura, H., and Dostie, J. (2011). Discovering genome regulation with 3C and 3C-related technologies. Biochimica et biophysica acta *1819*, 401-410.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Ferraiuolo, M.A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X.Q., Nadler, M., Blanchette, M., and Dostie, J. (2010). The three-dimensional architecture of Hox cluster silencing. Nucleic acids research *38*, 7472-7484.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J.*, et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.

Fischbach, N.A., Rozenfeld, S., Shen, W., Fong, S., Chrobak, D., Ginzinger, D., Kogan, S.C., Radhakrishnan, A., Le Beau, M.M., Largman, C.*, et al.* (2005). HOXB6 overexpression in murine bone marrow immortalizes a myelomonocytic precursor in vitro and causes hematopoietic stem cell expansion and acute myeloid leukemia in vivo. Blood *105*, 1456-1466.

Fitzsimmons, D., Lutz, R., Wheat, W., Chamberlin, H.M., and Hagman, J. (2001). Highly conserved amino acids in Pax and Ets proteins are required for DNA binding and ternary complex assembly. Nucleic acids research *29*, 4154-4165.

Floer, M., Wang, X., Prabhu, V., Berrozpe, G., Narayan, S., Spagna, D., Alvarez, D., Kendall, J., Krasnitz, A., Stepansky, A.*, et al.* (2010). A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. Cell *141*, 407-418.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490-493.

Galant, R., and Carroll, S.B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. Nature *415*, 910-913.

Galant, R., Walsh, C.M., and Carroll, S.B. (2002). Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. Development *129*, 3115-3126.

Gardner, K.E., Allis, C.D., and Strahl, B.D. (2011). Operating on chromatin, a colorful language where context matters. Journal of molecular biology *409*, 36-46.

Garvie, C.W., Hagman, J., and Wolberger, C. (2001). Structural studies of Ets-1/Pax5 complex formation on DNA. Molecular cell *8*, 1267-1276.

Gebelein, B., Culi, J., Ryoo, H.D., Zhang, W., and Mann, R.S. (2002). Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. Developmental cell *3*, 487-498.

Gebelein, B., and Mann, R.S. (2007). Compartmental modulation of abdominal Hox expression by engrailed and sloppy-paired patterns the fly ectoderm. Dev Biol *308*, 593-605.

Gebelein, B., McKay, D.J., and Mann, R.S. (2004). Direct integration of Hox and segmentation gene inputs during Drosophila development. Nature *431*, 653-659.

Gehring, W.J. (1987). Homeo boxes in the study of development. Science *236*, 1245-1252.

Georges, A.B., Benayoun, B.A., Caburet, S., and Veitia, R.A. (2010). Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? FASEB J *24*, 346-356.

Gibson, G., Schier, A., LeMotte, P., and Gehring, W.J. (1990). The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. Cell *62*, 1087-1103.

Gonzalez-Reyes, A., and Morata, G. (1990). The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interactions with other homeotic products. Cell *61*, 515-522.

Gonzalez-Reyes, A., and Morata, G. (1991). Organization of the Drosophila head as revealed by the ectopic expression of the Ultrabithorax product. Development *113*, 1459-1471.

Gonzalez-Reyes, A., Urquia, N., Gehring, W.J., Struhl, G., and Morata, G. (1990). Are cross-regulatory interactions between homoeotic genes functionally significant? Nature *344*, 78-80.

Gordan, R., Hartemink, A.J., and Bulyk, M.L. (2009). Distinguishing direct versus indirect transcription factor-DNA interactions. Genome research *19*, 2090-2100.

Green, N.C., Rambaldi, I., Teakles, J., and Featherstone, M.S. (1998). A conserved C-terminal domain in PBX increases DNA binding by the PBX homeodomain and is not a primary site of contact for the YPWM motif of HOXA1. J Biol Chem *273*, 13273-13279.

Guertin, M.J., and Lis, J.T. (2010). Chromatin landscape dictates HSF binding to target DNA elements. PLoS genetics *6*, e1001114.

Hafen, E., Levine, M., and Gehring, W.J. (1984). Regulation of Antennapedia transcript distribution by the bithorax complex in Drosophila. Nature *307*, 287-289.

Hakim, O., Sung, M.H., Voss, T.C., Splinter, E., John, S., Sabo, P.J., Thurman, R.E., Stamatoyannopoulos, J.A., de Laat, W., and Hager, G.L. (2011). Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. Genome research *21*, 697-706.

Harding, K., Wedeen, C., McGinnis, W., and Levine, M. (1985). Spatially regulated expression of homeotic genes in Drosophila. Science *229*, 1236-1242.

Harrison, M.M., Li, X.Y., Kaplan, T., Botchan, M.R., and Eisen, M.B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS genetics *7*, e1002266.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nature genetics *42*, 343-347.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nature genetics *43*, 414-420.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics *39*, 311-318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Heldring, N., Isaacs, G.D., Diehl, A.G., Sun, M., Cheung, E., Ranish, J.A., and Kraus, W.L. (2011). Multiple sequence-specific DNA-binding proteins mediate estrogen receptor signaling through a tethering pathway. Mol Endocrinol *25*, 564-574.

Hersh, B.M., Nelson, C.E., Stoll, S.J., Norton, J.E., Albert, T.J., and Carroll, S.B. (2007). The UBX-regulated network in the haltere imaginal disc of D. melanogaster. Dev Biol *302*, 717-727.

Hittinger, C.T., Stern, D.L., and Carroll, S.B. (2005). Pleiotropic functions of a conserved insect-specific Hox peptide motif. Development *132*, 5261-5270.

Hollenhorst, P.C., McIntosh, L.P., and Graves, B.J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. Annual review of biochemistry *80*, 437-471.

Hong, J.W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. Science *321*, 1314.

Hughes, C.L., and Kaufman, T.C. (2002). Hox genes and the evolution of the arthropod body plan. Evolution & development *4*, 459-499.

Iacovino, M., Hernandez, C., Xu, Z., Bajwa, G., Prather, M., and Kyba, M. (2009). A conserved role for Hox paralog group 4 in regulation of hematopoietic progenitors. Stem cells and development *18*, 783-792.

Istrail, S., and Davidson, E.H. (2005). Logic functions of the genomic cis-regulatory code. Proceedings of the National Academy of Sciences of the United States of America *102*, 4954-4959.

Jansen, A., van der Zande, E., Meert, W., Fink, G.R., and Verstrepen, K.J. (2012). Distal chromatin structure influences local nucleosome positions and gene expression. Nucleic acids research.

Jauch, R., Aksoy, I., Hutchins, A.P., Ng, C.K., Tian, X.F., Chen, J., Palasingam, P., Robson, P., Stanton, L.W., and Kolatkar, P.R. (2011). Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. Stem Cells *29*, 940-951.

Jegalian, B.G., and De Robertis, E.M. (1992). Homeotic transformations in the mouse induced by overexpression of a human Hox3.3 transgene. Cell *71*, 901-910.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

John, S., Sabo, P.J., Thurman, R.E., Sung, M.H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature genetics *43*, 264-268.

Jones, S.A., Shim, S.H., He, J., and Zhuang, X. (2011). Fast, three-dimensional super-resolution imaging of live cells. Nature methods *8*, 499-508.

Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Kong, S.L., Ukil, L., Pan, Y.F., Li, G., Lim, M., Thomsen, J.S.*, et al.* (2010). Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. Molecular systems biology *6*, 456.

Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell *131*, 530-543.

Joshi, R., Sun, L., and Mann, R. (2010). Dissecting the functional specificities of two Hox proteins. Genes & development *24*, 1533-1545.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell *148*, 473-486.

Kaplan, N., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010a). Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. Genome biology *11*, 140.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010b). Nucleosome sequence preferences influence in vivo nucleosome organization. Nature structural & molecular biology *17*, 918-920; author reply 920-912.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kaplan, T., Li, X.Y., Sabo, P.J., Thomas, S., Stamatoyannopoulos, J.A., Biggin, M.D., and Eisen, M.B. (2011). Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. PLoS genetics *7*, e1001290.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480-485.

Khoueiry, P., Rothbacher, U., Ohtsuka, Y., Daian, F., Frangulian, E., Roure, A., Dubchak, I., and Lemaire, P. (2010). A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites. Current biology : CB *20*, 792-802.

Knoepfler, P.S., Bergstrom, D.A., Uetsuki, T., Dac-Korytko, I., Sun, Y.H., Wright, W.E., Tapscott, S.J., and Kamps, M.P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic acids research *27*, 3752-3761.

Knoepfler, P.S., Sykes, D.B., Pasillas, M., and Kamps, M.P. (2001). HoxB8 requires its Pbx-interaction motif to block differentiation of primary myeloid progenitors and of most cell line models of myeloid differentiation. Oncogene *20*, 5440-5448.

Kondoh, H., and Kamachi, Y. (2010). SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. The international journal of biochemistry & cell biology *42*, 391-399.

Krivega, I., and Dean, A. (2011). Enhancer and promoter interactions-long distance calls. Current opinion in genetics & development.

Kulkarni, M.M., and Arnosti, D.N. (2003). Information display by transcriptional enhancers. Development *130*, 6569-6575.

Kuziora, M.A., and McGinnis, W. (1988). Autoregulation of a Drosophila homeotic selector gene. Cell *55*, 477-485.

LaRonde-LeBlanc, N.A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes & development *17*, 2060-2072.

Lee, J.S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. Cell *142*, 682-685.

Lelli, K.M., Noro, B., and Mann, R.S. (2011). Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. Proceedings of the National Academy of Sciences of the United States of America *108*, 21122-21127.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current opinion in genetics & development *21*, 175-186.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J.*, et al.* (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84-98.

Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L.*, et al.* (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS biology *6*, e27.

Li, X.Y., Thomas, S., Sabo, P.J., Eisen, M.B., Stamatoyannopoulos, J.A., and Biggin, M.D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome biology *12*, R34.

Li-Kroeger, D., Witt, L.M., Grimes, H.L., Cook, T.A., and Gebelein, B. (2008). Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the Drosophila PNS. Developmental cell *15*, 298-308.

Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400-403.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, L., and McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. Genes & development *6*, 1071-1081.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell *126*, 403-413.

Lomvardas, S., and Thanos, D. (2002). Modifying gene expression programs by altering core promoter chromatin architecture. Cell *110*, 261-271.

Ludwig, M.Z., Manu, Kittler, R., White, K.P., and Kreitman, M. (2011). Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS genetics *7*, e1002364.

Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. PLoS biology *3*, e93.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958-970.

MacArthur, S., Li, X.Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keranen, S.V.*, et al.* (2009). Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome biology *10*, R80.

MacQuarrie, K.L., Fong, A.P., Morse, R.H., and Tapscott, S.J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. Trends in genetics : TIG *27*, 141-148.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends in genetics : TIG *27*, 465-474.

Mann, R.S., and Chan, S.K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends in genetics : TIG *12*, 258-262.

Mann, R.S., and Hogness, D.S. (1990). Functional dissection of Ultrabithorax proteins in D. melanogaster. Cell *60*, 597-610.

Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. Curr Top Dev Biol *88*, 63-101.

Mao, C., Brown, C.R., Griesenbeck, J., and Boeger, H. (2011). Occlusion of regulatory sequences by promoter nucleosomes in vivo. PloS one *6*, e17521.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283-302.

Medina-Martinez, O., and Ramirez-Solis, R. (2003). In vivo mutagenesis of the Hoxb8 hexapeptide domain leads to dominant homeotic transformations that mimic the loss-of-function mutations in genes of the Hoxb cluster. Dev Biol *264*, 77-90.

Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407-410.

Merabet, S., Kambris, Z., Capovilla, M., Berenger, H., Pradel, J., and Graba, Y. (2003). The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. Developmental cell *4*, 761-768.

Merabet, S., Litim-Mecheri, I., Karlsson, D., Dixit, R., Saadaoui, M., Monier, B., Brun, C., Thor, S., Vijayraghavan, K., Perrin, L.*, et al.* (2011). Insights into Hox protein function from a large scale combinatorial analysis of protein domains. PLoS genetics *7*, e1002302.

Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., and Graba, Y. (2007). A unique Extradenticle recruitment mode in the Drosophila Hox protein Ultrabithorax. Proceedings of the National Academy of Sciences of the United States of America *104*, 16946-16951.

Merabet, S., Sambrani, N., Pradel, J., and Graba, Y. (2010). Regulation of Hox activity: insights from protein motifs. Adv Exp Med Biol *689*, 3-16.

Miller, J.A., and Widom, J. (2003). Collaborative competition mechanism for gene activation in vivo. Molecular and cellular biology *23*, 1623-1632.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. Proceedings of the National Academy of Sciences of the United States of America *107*, 22534-22539.

Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. (2001). Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. Journal of molecular biology *314*, 495-506.

Moens, C.B., and Selleri, L. (2006). Hox cofactors in vertebrate development. Dev Biol *291*, 193-206.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. Cell *147*, 1132-1145.

Moorman, C., Sun, L.V., Wang, J., de Wit, E., Talhout, W., Ward, L.D., Greil, F., Lu, X.J., White, K.P., Bussemaker, H.J.*, et al.* (2006). Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America *103*, 12027-12032.

Moretti, R., and Ansari, A.Z. (2008). Expanding the specificity of DNA targeting by harnessing cooperative assembly. Biochimie *90*, 1015-1025.

Moyle-Heyrman, G., Tims, H.S., and Widom, J. (2011). Structural constraints in collaborative competition of transcription factors against the nucleosome. Journal of molecular biology *412*, 634-646.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R.*, et al.* (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527-531.

Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.*, et al.* (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS genetics *6*, e1000814.

Ng, C.K., Li, N.X., Chee, S., Prabhakar, S., Kolatkar, P.R., and Jauch, R. (2012). Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. Nucleic acids research.

Nien, C.Y., Liang, H.L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J.R., and Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS genetics *7*, e1002339.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. Science *334*, 222-225.

Noro, B., Culi, J., McKay, D.J., Zhang, W., and Mann, R.S. (2006). Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by homothorax. Genes & development *20*, 1636-1650.

Noro, B., Lelli, K., Sun, L., and Mann, R.S. (2011). Competition for cofactor-dependent DNA binding underlies Hox phenotypic suppression. Genes & development *25*, 2327-2332.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell *133*, 1277-1289.

O'Connor, M.B., Binari, R., Perkins, L.A., and Bender, W. (1988). Alternative RNA products from the Ultrabithorax domain of the bithorax complex. The EMBO journal *7*, 435-445.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W.*, et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nature genetics *36*, 1065-1071.

Pan, Y., and Nussinov, R. (2011). The role of response elements organization in transcription factor selectivity: the IFN-beta enhanceosome example. PLoS computational biology *7*, e1002077.

Pan, Y., Tsai, C.J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. Trends in genetics : TIG *26*, 75-83.

Panne, D. (2008). The enhanceosome. Current opinion in structural biology *18*, 236-242.

Panne, D., Maniatis, T., and Harrison, S.C. (2004). Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. The EMBO journal *23*, 4384-4393.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. Cell *129*, 1111-1123.

Papadopoulos, D.K., Resendez-Perez, D., Cardenas-Chavez, D.L., Villanueva-Segura, K., Canales-del-Castillo, R., Felix, D.A., Funfschilling, R., and Gehring, W.J. (2011). Functional synthetic Antennapedia genes and the dual roles of YPWM motif and linker size in transcriptional activation and repression. Proceedings of the National Academy of Sciences of the United States of America *108*, 11959-11964.

Papadopoulos, D.K., Vukojevic, V., Adachi, Y., Terenius, L., Rigler, R., and Gehring, W.J. (2010). Function and specificity of synthetic Hox transcription factors in vivo. Proceedings of the National Academy of Sciences of the United States of America *107*, 4087-4092.

Parker, D.S., White, M.A., Ramos, A.I., Cohen, B.A., and Barolo, S. (2011). The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. Science signaling *4*, ra38.

Peifer, M., and Wieschaus, E. (1990). Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. Genes & development *4*, 1209-1223.

Peltenburg, L.T., and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. The EMBO journal *15*, 3385-3393.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Current biology : CB *20*, 1562-1567.

Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. Journal of molecular biology *258*, 800-812.

Rando, O.J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. Current opinion in genetics & development.

Rando, O.J., and Chang, H.Y. (2009). Genome-wide views of chromatin structure. Annual review of biochemistry *78*, 245-271.

Rauskolb, C., Peifer, M., and Wieschaus, E. (1993). extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. Cell *74*, 1101-1112.

Rauskolb, C., and Wieschaus, E. (1994). Coordinate regulation of downstream genes by extradenticle and the homeotic selector proteins. The EMBO journal *13*, 3561-3569.

Razin, S.V., Gavrilov, A.A., Pichugin, A., Lipinski, M., Iarovaia, O.V., and Vassetzky, Y.S. (2011). Transcription factories in the context of the nuclear and genome organization. Nucleic acids research *39*, 9085-9092.

Remacle, S., Abbas, L., De Backer, O., Pacico, N., Gavalas, A., Gofflot, F., Picard, J.J., and Rezsohazy, R. (2004). Loss of function but no gain of function caused by amino acid substitutions in the hexapeptide of Hoxa1 in vivo. Molecular and cellular biology 24, 8567-8575.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., and Mann, R.S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. Cell *91*, 171-183.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annual review of biochemistry *79*, 233-269.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. Nature *461*, 1248-1253.

Ronshaugen, M., Biemar, F., Piel, J., Levine, M., and Lai, E.C. (2005). The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. Genes & development *19*, 2947-2952.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nature *415*, 914-917.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F*., et al.* (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787-1797.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature methods *3*, 793-795.

Ryoo, H.D., and Mann, R.S. (1999). The control of trunk Hox specificity and activity by Extradenticle. Genes & development *13*, 1704-1716.

Saadaoui, M., Merabet, S., Litim-Mecheri, I., Arbeille, E., Sambrani, N., Damen, W., Brena, C., Pradel, J., and Graba, Y. (2011). Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. Proceedings of the National Academy of Sciences of the United States of America *108*, 2276-2281.

Sanchez-Herrero, E., Vernos, I., Marco, R., and Morata, G. (1985). Genetic organization of Drosophila bithorax complex. Nature *313*, 108-113.

Satija, R., and Bradley, R.K. (2012). The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. Genome research.

Schnabel, C., Jacobs Y., and Cleary, M.L. (2000). HoxA9-mediated immortalization of myeloid pregenitors requires functional interactions with TALE cofactors Pbx and Meis. Oncogene *19,* 608-616.

Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816-818.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S*., et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nature genetics *42*, 53-61.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends in genetics : TIG *25*, 335-343.

Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A.V., and Zaret, K.S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes & development *23*, 804-809.

Sekiya, T., and Zaret, K.S. (2007). Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. Molecular cell *28*, 291-303.

Serandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R*., et al.* (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome research *21*, 555-565.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Shen, W., Chrobak, D., Krishnan, K., Lawrence, H.J., and Largman, C. (2004). HOXB6 protein is bound to CREB-binding protein and represses globin expression in a DNA binding-dependent, PBX interaction-independent process. The Journal of biological chemistry *279*, 39895-39904.

Shykind, B.M. (2005). Regulation of odorant receptors: one allele at a time. Human molecular genetics *14 Spec No 1*, R33-39.

Siggers, T., Duyzend, M.H., Reddy, J., Khan, S., and Bulyk, M.L. (2011). Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. Molecular systems biology *7*, 555.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nature reviews Molecular cell biology *10*, 697-708.

Singh, N.P., and Mishra, R.K. (2008). A double-edged sword to force posterior dominance of Hox genes. BioEssays : news and reviews in molecular, cellular and developmental biology *30*, 1058-1061.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J*., et al.* (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell *147*, 1270-1282.

Stevens, K.E., and Mann, R.S. (2007). A balance between two nuclear localization sequences and a nuclear export sequence governs extradenticle subcellular localization. Genetics *175*, 1625-1636.

Stobe, P., Stein, M.A., Habring-Muller, A., Bezdan, D., Fuchs, A.L., Hueber, S.D., Wu, H., and Lohmann, I. (2009). Multifactorial regulation of a hox target gene. PLoS genetics *5*, e1000412.

Struhl, G. (1983). Role of the esc+ gene product in ensuring the selective expression of segment-specific homeotic genes in Drosophila. J Embryol Exp Morphol *76*, 297-331.

Struhl, G., and White, R.A. (1985). Regulation of the Ultrabithorax gene of Drosophila by other bithorax complex genes. Cell *43*, 507-519.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Developmental cell *18*, 359-370.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nature reviews Molecular cell biology *11*, 264-275.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Thomas, S., Li, X.Y., Sabo, P.J., Sandstrom, R., Thurman, R.E., Canfield, T.K., Giste, E., Fisher, W., Hammonds, A., Celniker, S.E*., et al.* (2011). Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome biology *12*, R43.

Tims, H.S., Gurunathan, K., Levitus, M., and Widom, J. (2011). Dynamics of nucleosome invasion by DNA binding proteins. Journal of molecular biology *411*, 430-448.

Tolkunov, D., Zawadzki, K.A., Singer, C., Elfving, N., Morozov, A.V., and Broach, J.R. (2011). Chromatin remodelers clear nucleosomes from intrinsically unfavorable sites to establish nucleosome-depleted regions at promoters. Molecular biology of the cell *22*, 2106-2118.

Tootle, T.L., and Rebay, I. (2005). Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. BioEssays : news and reviews in molecular, cellular and developmental biology *27*, 285-298.

Tour, E., Hittinger, C.T., and McGinnis, W. (2005). Evolutionarily conserved domains required for activation and repression functions of the Drosophila Hox protein Ultrabithorax. Development *132*, 5271-5281.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Uhl, J.D., Cook, T.A., and Gebelein, B. (2010). Comparing anterior and posterior Hox complex formation reveals guidelines for predicting cis-regulatory elements. Dev Biol *343*, 154-166.

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M.E., Botas, J., and Cohen, S.M. (1992). Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less. Cell *71*, 437-450.

van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nature biotechnology *18*, 424-428.

Vaquerizas, J.M., Akhtar, A., and Luscombe, N.M. (2011). Large-scale nuclear architecture and transcriptional control. Sub-cellular biochemistry *52*, 279-295.

Verger, A., and Duterque-Coquillaud, M. (2002). When Ets transcription factors meet their partners. BioEssays : news and reviews in molecular, cellular and developmental biology *24*, 362-370.

Vitobello, A., Ferretti, E., Lampe, X., Vilain, N., Ducret, S., Ori, M., Spetz, J.F., Selleri, L., and Rijli, F.M. (2011). Hox and Pbx factors control retinoic acid synthesis during hindbrain segmentation. Developmental cell *20*, 469-482.

Voss, T.C., Schiltz, R.L., Sung, M.H., Yen, P.M., Stamatoyannopoulos, J.A., Biddie, S.C., Johnson, T.A., Miranda, T.B., John, S., and Hager, G.L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell *146*, 544-554.

Wieschaus, E., Nusslein-Volhard C. (1986). Looking at embryos. (Oxford: IRL Press Limited).

Williams, T.M., Williams, M.E., and Innis, J.W. (2005). Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. Dev Biol *277*, 457-471.

Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends in genetics : TIG *25*, 434-440.

Xie, B., Charlton-Perkins, M., McDonald, E., Gebelein, B., and Cook, T. (2007). Senseless functions as a molecular switch for color photoreceptor differentiation in Drosophila. Development *134*, 4243-4253.

Yaron, Y., McAdara, J.K., Lynch, M., Hughes, E., and Gasson, J.C. (2001). Identification of novel functional regions important for the activity of HOXB7 in mammalian cells. J Immunol *166*, 5058-5067.

Yekta, S., Tabin, C.J., and Bartel, D.P. (2008). MicroRNAs in the Hox network: an apparent link to posterior prevalence. Nature reviews Genetics *9*, 789-796.

You, J.S., Kelly, T.K., De Carvalho, D.D., Taberlay, P.C., Liang, G., and Jones, P.A. (2011). OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proceedings of the National Academy of Sciences of the United States of America *108*, 14497-14502.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & development *25*, 2227-2241.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nature reviews Genetics *12*, 7-18.

Zhou, X., and O'Shea, E.K. (2011). Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. Molecular cell *42*, 826-836.

Zhu, X., Ahmad, S.M., Aboukhalil, A., Busser, B.W., Kim, Y., Tansey, T.R., Haimovich, A., Jeffries, N., Bulyk, M.L., and Michelson, A.M. (2012). Differential regulation of mesodermal gene expression by Drosophila cell type-specific Forkhead transcription factors. Development *139*, 1457-1466.

Zorn, C., Cremer, C., Cremer, T., and Zimmer, J. (1979). Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. Experimental cell research *124*, 111-119.

Zuleger, N., Robson, M.I., and Schirmer, E.C. (2011). The nuclear envelope as a chromatin organizer. Nucleus *2*, 339-349.

Appendix 1.

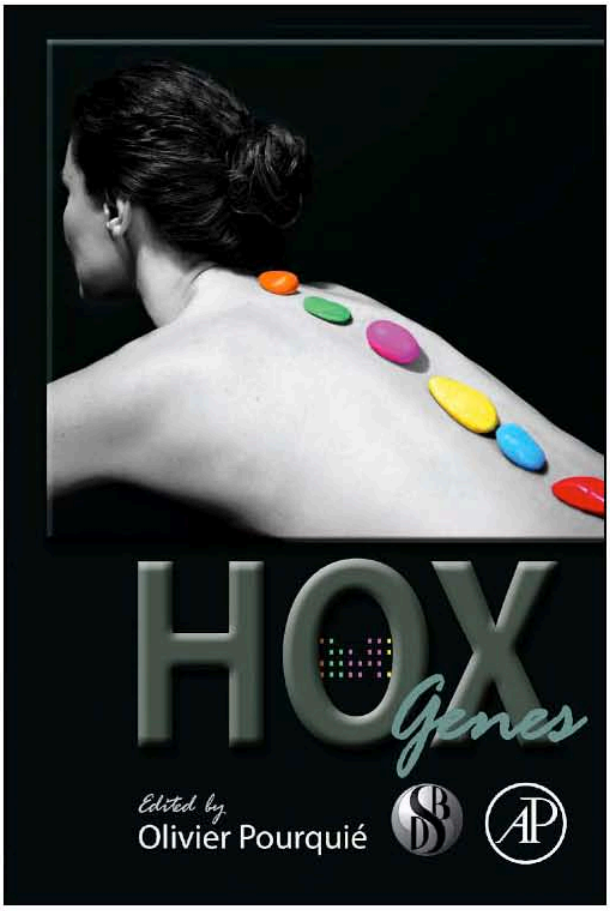
Hox Specificity: Unique Roles for Cofactors and Collaborators.

Richard S. Mann, Katherine M. Lelli, and Rohit Joshi, Hox Specificity: Unique Roles for Cofactors and Collaborators. In Olivier Pourquie, editor: *Current Topics in Developmental Biology*, Vol. 88, Burlington: Academic Press, 2009, pp. 63-101.

I have included this chapter to provide the reader with a more detailed background into Hox-cofactor interactions. My specific contributions included: creating the logos in figures 3.2, generating the images of the structures presented in Figures 3.4 and 3.5 and curating all of the known Hox targets to generate Table 3.1. I was also in charge of the referencing.

Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.

This chapter was originally published in the book *Current Topics in Developmental Biology,* Vol. 88, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:
http://www.elsevier.com/locate/permissionusematerial

From: Richard S. Mann, Katherine M. Lelli, and Rohit Joshi, Hox Specificity: Unique Roles for Cofactors and Collaborators. In Olivier Pourquie, editor: Current Topics in Developmental Biology, Vol. 88, Burlington: Academic Press, 2009, pp. 63-101.
ISBN: 978-0-12-374529-3
© Copyright 2009 Elsevier Inc.
Academic Press.

Author's personal copy

CHAPTER THREE

Hox Specificity: Unique Roles for Cofactors and Collaborators

Richard S. Mann,* Katherine M. Lelli,† *and* Rohit Joshi*

Contents

Abstract

Hox proteins are well known for executing highly specific functions *in vivo*, but our understanding of the molecular mechanisms underlying gene regulation by these fascinating proteins has lagged behind. The premise of this review is that an understanding of gene regulation—by any transcription factor—requires the dissection of the *cis*-regulatory elements that they act upon. With this goal in mind, we review the concepts and ideas regarding gene regulation by Hox proteins and apply them to a curated list of directly regulated Hox *cis*-regulatory elements that have been validated in the literature. Our analysis of the Hox-binding sites within these elements suggests several emerging generalizations. We distinguish between Hox cofactors, proteins that bind DNA cooperatively with Hox proteins and thereby help with DNA-binding site selection, and Hox collaborators, proteins that bind in parallel to Hox-targeted *cis*-regulatory elements and dictate the sign and strength of gene regulation. Finally, we summarize insights that come from examining five X-ray crystal structures of

* Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA
† Department of Genetics and Development, Columbia University, New York, NY, USA

Current Topics in Developmental Biology, Volume 88
ISSN 0070-2153, DOI: 10.1016/S0070-2153(09)88003-4
© 2009 Elsevier Inc.
All rights reserved.

Author's personal copy

Hox-cofactor-DNA complexes. Together, these analyses reveal an enormous amount of flexibility into how Hox proteins function to regulate gene expression, perhaps providing an explanation for why these factors have been central players in the evolution of morphological diversity in the animal kingdom.

1. An Introduction to the Problem

Hox proteins are homeodomain-containing transcription factors that have the capacity to carry out exquisitely precise functions *in vivo* that are critical for many aspects of animal morphogenesis. Most typically, each *Hox* gene is expressed in a subset of the anterior-posterior (AP) body axis, where it specifies cellular and tissue identities. Famous examples of the power that *Hox* genes have to sculpt animal morphogenesis include the antenna-to-leg transformation caused by the *Antennapedia* (*Antp*) mutation in *Drosophila* and several polydactyly syndromes in humans (Goodman, 2002; Lewis, 1978; Randazzo *et al.*, 1991). These types of phenotypes have been a long-standing source of fascination for both biologists and lovers of science fiction.

An important and long-debated question for Hox biologists has been how these proteins achieve this apparently high degree of *in vivo* specificity. In this review, we summarize ideas and recent data bearing on the question of Hox specificity, with a special emphasis on what can be learned by analyzing native *cis*-regulatory elements that are directly bound and regulated by Hox proteins. Excellent reviews discussing the range of Hox target genes that have been identified using genome-wide and traditional approaches complement the emphasis of this chapter (Hueber and Lohmann, 2008; Pearson *et al.*, 2005).

When *Hox* genes were first cloned and shown to encode homeodomain-containing proteins (Akam, 1989; Regulski *et al.*, 1985), researchers initially speculated that Hox proteins would bind and regulate the correct subset of target genes according to the DNA recognition properties of their homeodomains. However, early work from a number of labs quickly established that homeodomains were not likely to be up to the task of dictating Hox-DNA-binding specificities on their own (Affolter *et al.*, 1990; Desplan *et al.*, 1988; Ekker *et al.*, 1991, 1994; Hoey and Levine, 1988). Indeed, homeodomains, particularly the subset present in the Hox protein family, all bind to a very similar set of "AT"-rich DNA-binding sites, raising the fundamental question of how specificity is achieved. In addition to this classical problem of degenerate binding site recognition, experiments using chimeric Hox proteins—where bits of one Hox protein were replaced with homologous bits of another—highlight an additional complication. As expected, specific Hox functions required the homeodomain. In some cases, however, specificity also required nonhomeodomain

Author's personal copy

residues, in particular, those that lie immediately N- or C-terminal to the homeodomain (Chan *et al.*, 1994; Dessain *et al.*, 1992; Furukubo-Tokunaga *et al.*, 1993; Gibson *et al.*, 1990; Kuziora and McGinnis, 1989, 1990; Lin and McGinnis, 1992; Mann and Hogness, 1990; Zhao and Potter, 2001, 2002). How these nonhomeodomain residues may impact Hox specificity is just now coming into focus.

Because they are DNA-binding transcription factors, it is probably a safe bet that Hox proteins carry out the majority of their functions by binding to *cis*-regulatory elements (although alternative mechanisms have been proposed; Plaza *et al.*, 2008). Because eukaryotic transcription is governed by *cis*-regulatory elements that typically integrate multiple inputs, each Hox-targeted element is likely to have binding sites for many transcription factors. Therefore, to understand how Hox proteins ultimately function to control target gene expression, it is necessary to consider two broad questions. First, how do Hox proteins recognize their DNA-binding sites and second, how do they interact with other transcriptional inputs that feed into the same *cis*-regulatory element? We suggest that it is helpful to break the problem of Hox specificity down into two conceptually separable steps (Fig. 3.1). In the first step, the question can be rephrased to ask: How do Hox proteins find the right DNA-binding sites *in vivo*? Many examples exist in the literature suggesting that Hox proteins solve this initial "DNA-binding specificity" step in multiple ways. As will be explored more fully below, one solution is by the use of cooperatively binding cofactors such as Extradenticle (Exd), Pbx, Homothorax (Hth), and Meis that increase Hox-DNA-binding specificities (previously reviewed by Mann and Affolter, 1998; Mann and Chan, 1996; Moens and Selleri, 2006). However, it is also increasingly clear that Hox proteins regulate many genes without the help of these cofactors. In the second step, the question is: Once bound, how do Hox proteins orchestrate a transcriptional response? As the same Hox protein can activate some target genes, and repress others, it is clear that this "activity regulation" step is also critical for how Hox proteins execute their *in vivo* functions. In fact, as will be described more fully below, there is now good evidence for both of the steps outlined in Fig. 3.1 playing critical roles in Hox specificity.

2. Too Many Binding Sites, Not Enough Specificity

Because all Hox proteins have a homeodomain, understanding how Hox proteins recognize their DNA-binding sites *in vivo* certainly depends, at least in part, on how this 60 amino acid domain recognizes DNA sequences. The basic DNA recognition principles for homeodomains

Author's personal copy

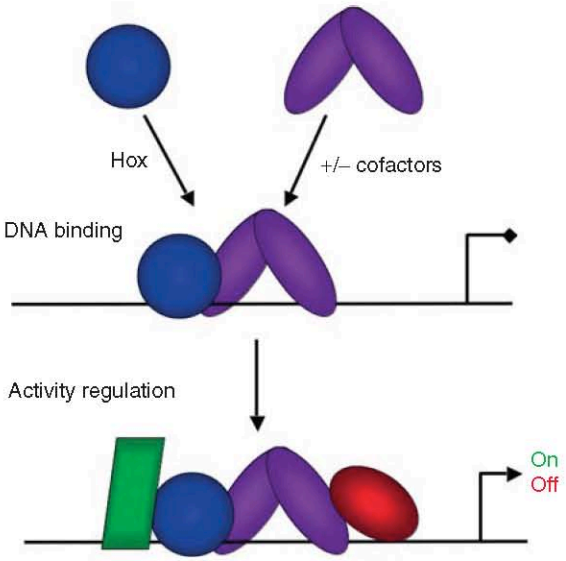


Figure 3.1 Two contributing steps in Hox specificity. In principle, Hox specificity can be broken down into two separate steps. The first step is DNA binding by Hox proteins, which can occur either with or without cooperatively binding cofactors. The second step involves the recruitment of additional factors, Hox collaborators, to the *cis*-regulatory element. The recruitment of these factors may depend on contacts between them and the DNA and/or protein-protein contacts between them and the Hox-cofactor complex. It is the recruitment of these collaborators, which we suggest depends on the architecture of the entire *cis*-regulatory element (including the details of the Hox-binding site) that ultimately determines the sign of the transcriptional regulation.

were established from biochemical and structural studies (reviewed previously by Gehring *et al.*, 1994). These studies show that all homeodomains fold into a bundle of three alpha-helices and an unstructured "N-terminal" arm. DNA contacts are formed primarily by residues 47, 50, 51, and 54 in the third alpha-helix (the so-called recognition helix) and by an arginine in position 5 of the N-terminal arm. While these studies provided a high resolution view of how homeodomains generally bind to DNA, they did not provide much insight into the problem of Hox specificity for three reasons. First, nearly all Hox homeodomains, even those with very disparate *in vivo* functions, have the same residues in the DNA-contacting residues visualized in these structures (Mann, 1995). Second, although nonhomeodomain residues were known to play a role in Hox specificity from studies of chimeric Hox proteins (see above), these domains were not present in any of the initial structural studies. Third, the DNA sequences used in these early structural studies were not *in vivo* binding sites. Instead, these structures used high-affinity consensus sites that would not be expected to reveal insights into homeodomain specificity.

Author's personal copy

Another limitation in the early studies on homeodomain-DNA recognition was that only a small subset of homeodomain proteins were studied. Thanks to the advent of new, powerful methodologies, two large-scale studies have recently defined the individual DNA-binding specificities for nearly all homeodomains, including the subset present in the Hox proteins. One group used a bacterial one-hybrid approach (B1H) to analyze the DNA-binding preferences for all of the homeodomains encoded in the *Drosophila melanogaster* genome (Noyes *et al.*, 2008) while the second group used an entirely *in vitro* platform, protein-binding microarrays (PBMs), to answer the same question for all mouse homeodomains (Berger *et al.*, 2008). Although there are pros and cons for each approach (Affolter *et al.*, 2008), these studies confirmed that Hox homeodomains like to bind "AT"-rich DNA sequences (Fig. 3.2). In particular, the so-called "Antennapedia (Antp) Group" of homeodomains, which includes all Hox homeodomains except for those of the Abdominal-B (Abd-B) class, like to bind the sequence TAAT[t/g][a/g]. There are 87,307 copies of the sequence TAATTA and 86,201 copies of the sequence TAATGA in the *D. melanogaster* genome, each more than five times the number of annotated protein-coding genes. Clearly, the presence of TAAT[t/g][a/g] cannot be sufficient information for Hox regulation. Moreover, as elegantly illustrated by the homeodomain-binding site survey studies (Berger *et al.*, 2008; Noyes *et al.*, 2008), TAAT[t/g][a/g] is readily bound by most Hox homeodomains, as well as many non-Hox homeodomains. Therefore, this and related binding sites cannot be sufficient to distinguish between Hox family members that carry out distinct functions *in vivo*.

3. How Specific Do Hox Proteins Need to be?

Hox biologists can readily point to highly specific functions that are uniquely specified by individual Hox proteins. For example, in *Drosophila*, only the *Hox* gene *Sex combs reduced* (*Scr*) can orchestrate the development of a salivary gland, presumably by regulating a network of salivary gland-promoting genes (Bradley *et al.*, 2001). The flip view that multiple Hox proteins probably share many targets is typically given less attention. We believe this discussion is highly relevant to how one thinks about Hox specificity, because it may be that only a subset of Hox targets for any particular Hox protein need be highly specific.

One example of a morphological process that is controlled by multiple *Hox* genes is appendage development in *Drosophila*. Leg development is limited to the thoracic segments due to repression of *Distalless* (*Dll*) by the abdominal *Hox* genes, *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*) (Estrada and Sánchez-Herrero, 2001; Vachon *et al.*, 1992).

Author's personal copy

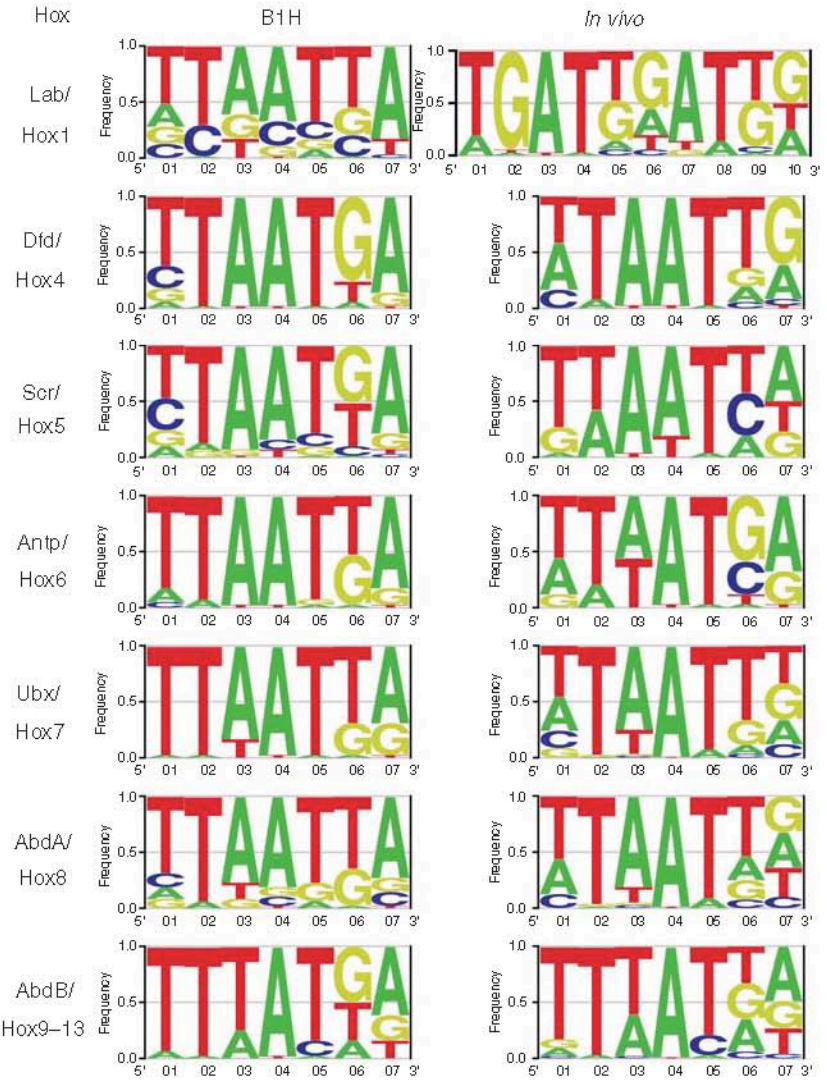


Figure 3.2 Comparison of *in vitro* and *in vivo* Hox-binding site preferences. Shown are LOGO diagrams summarizing Hox-binding site preferences for most paralogs. The column on the left lists the LOGOs generated using the binding sites identified by the bacterial 1-hybrid (B1H) method (Noyes *et al.*, 2008). The column on the right lists the LOGOs generated using the *in vivo* binding sites in Table 3.1. To generate these LOGOs, we used CONSENSUS (as part of Target Explorer; http://luna.bioc.columbia.edu/Target_Explorer/) to generate position weight matrices (PWMs). PWMs that maximized alignment of an "AT" sequence were converted to Transfac format using the phiSITE conversion server (http://www.phisite.org/main/index.php?nav=home). enoLOGOS (http://chianti.ucsd.edu/cgi-bin/enologos/enologos.cgi/) was then used to generate the LOGOs using nucleotide frequency for the *Y*-axis.

Author's personal copy

Moreover, *Ubx* and *abd-A* work, at least in part, through common binding sites in a *Dll cis*-regulatory element (Gebelein *et al.*, 2002, 2004). Thus, *Dll* is a shared target for the abdominal *Hox* genes. Similarly, the antenna-specifying gene *homothorax* (*hth*) can be repressed by all of the trunk Hox proteins, suggesting that *hth* is also a shared target (Casares and Mann, 1998, 2001; Yao *et al.*, 1999). A third example is the *Drosophila* head-promoting gene, *optix*, which is repressed by the trunk *Hox* genes and activated by the more anterior (head) *Hox* genes (Coiffier *et al.*, 2008). For *hth* and *optix*, a limitation in this conclusion is that the Hox-binding sites in these genes (assuming the regulation is direct) have not yet been identified. Therefore, although these genes are clearly shared *Hox* targets by genetic criteria, it is possible that different Hox proteins use different binding sites within these genes to regulate their expression. Nevertheless, these and other examples (Greig and Akam, 1995; Hirth *et al.*, 2001) support the idea that not all Hox functions need to be paralog-specific. If true, it follows that many bona fide Hox-binding sites may not need to discriminate between different Hox proteins.

How many Hox targets are shared and how many are paralog-specific? Although the field may be getting closer to a definitive answer to this question, by applying ChIP-chip and/or ChIP-seq methodologies to Hox proteins, the currently available data provide an interesting estimate. Using overexpression of Hoxc8 in mouse embryo fibroblasts, the expression levels of 34 genes were found to change by twofold or more (Lei *et al.*, 2005). This relatively small number of potential Hoxc8 targets contrasts from the much larger numbers of regulated genes identified in whole embryo expression profiling experiments following the uniform (and ectopic) expression of individual Hox proteins during *Drosophila* embryogenesis (Hueber *et al.*, 2007). An important strength of these experiments is that the global transcriptional response to multiple Hox proteins was analyzed in parallel, using the same experimental conditions. Remarkably, of the ~1500 genes (about 10% of all *Drosophila* genes) whose expression changed significantly in response to ectopic Deformed (Dfd), Scr, Antp, Ubx, Abd-A, or Abd-B expression, more than two-thirds (~69%) were regulated by only one of these six Hox proteins. About one-third (~30%) of all Hox-responsive target genes responded to multiple Hox proteins, while only ~1% responded to all six of these Hox proteins. There are, however, a few caveats to these experiments. For one, they measured responses to ectopic

The number of binding sites used to generate each LOGO was as follows: Labial: 31 (B1H), 17 (*in vivo*); Dfd: 24 (B1H), 17 (*in vivo*); Scr: 34 (B1H), 12 (*in vivo*); Antp: 19 (B1H), 16 (*in vivo*); Ubx: 20 (B1H), 57 (*in vivo*; the resulting LOGO was only subtly affected if the 30 sites from the Antp-P2 element were omitted); Abd-A: 23 (B1H), 39 (*in vivo*); and AbdB: 21 (B1H), 49 (*in vivo*).

Author's personal copy

Hox expression, which may not always reflect accurate gene regulation in their native expression domains. Second, these experiments did not distinguish between tissue-specific responses and third, they could not unambiguously discriminate between direct and indirect effects. Nevertheless, the results from this study are remarkable because they suggest that many, and perhaps the majority of Hox target genes are paralog-specific. However, the results also support the view that a significant number of genes are targeted by multiple Hox proteins, again raising the possibility that not all Hox-binding sites need to discriminate between Hox proteins.

Developmental context is another issue that should be considered when thinking about Hox specificity. In *Drosophila*, for example, each embryonic segment is built using the same reiterated set of signaling pathways that provide them with a common coordinate system, known as positional information. Once this "ground plan" is established, the non-*Hox* regulatory inputs into a gene are largely the same from segment to segment. One way to think about the *Hox* genes is that they impose identity information on top of this developmental ground plan, thus providing each segment its unique characteristics. Because the other regulatory inputs are largely equivalent, a gene that is specifically expressed (or repressed) in a small subset of embryonic segments is likely to be a paralog-specific Hox target gene. For example, for salivary glands to form only in the first thoracic (T1) segment, the *Hox* gene *Scr* must activate the salivary gland program, including its directly activated target gene *forkhead* (*fkh*), in a paralog-specific manner; other Hox proteins cannot activate this target in other segments (Bradley *et al.*, 2001). Similarly, while *Dll* is repressed by abdominal Hox proteins, the thoracic Hox proteins Scr and Antp must be permissive for *Dll* expression (Gebelein *et al.*, 2002). Repression by the abdominal Hox proteins therefore must have specificity. However, the same is not true for tissues where the ground plan, and thus the other regulatory inputs, is different. For example, the *Hox* gene *Ubx* is expressed in all cells of the developing haltere (a balancing organ used during flight) of the fly, where it regulates the expression of many genes (Crickmore and Mann, 2007, 2008; Lewis, 1978; Weatherbee and Carroll, 1999). Other than the developing wing, where *Hox* genes are not expressed for most of development, no other tissue in the fly has the same ground plan as the haltere. Therefore, the *cis*-regulatory elements used by *Ubx* to regulate genes in the haltere may not need to be highly selective for Ubx: other Hox proteins never have the opportunity to regulate these genes in the haltere/wing tissue simply because they are not expressed there. Confirmation of this idea comes from the finding that other Hox proteins, when expressed in the wing, can result in haltere-like phenotypes and mimic Ubx-like regulation (Casares *et al.*, 1996; R. S. Mann and M. Crickmore, unpublished observations). Similarly, although both Ubx and Abd-A have the potential to induce gonad development, the job is normally carried out by Abd-A

Author's personal copy

simply because it is the only Hox protein that is expressed in the correct set of mesodermal progenitor cells (Greig and Akam, 1995). Finally, another remarkable example of functional redundancy among different Hox paralogs is that all of the *Drosophila* Hox proteins, except for Abd-B, have the ability to replace Labial in the specification of the tritocerebral neuromere in the fly's brain (Hirth *et al.*, 2001). Like Ubx in the haltere, the ground plan for this portion of the brain may be sufficiently unique so that it does not require exquisite Hox specificity, accounting for why nearly all Hox paralogs can, at least to some degree, carry out the same functions as Labial.

In summary, when thinking about what Hox-bindings sites may look like based on these considerations, we suggest that it is useful to distinguish between three types of Hox target genes (Fig. 3.3) (1) those that must be highly specific for one Hox paralog ("paralog-specific"; e.g., Scr → *fkh*), (2) those that are shared by a subset of Hox proteins ("semi-paralog-specific"; e.g., Ubx, Abd-A, and Abd-B ⊣ *Dll*), and (3) those that are regulated by most or all *Hox* genes ("general"; e.g., all of the trunk *Hox* genes ⊣ *optix*). In addition, we argue that for some Hox functions that have the appearance of paralog specificity (e.g., Ubx dictating haltere instead of wing fates), the Hox-binding sites, themselves, do not need to be paralog-specific as long as the developmental context is sufficient to specify a unique regulatory environment.

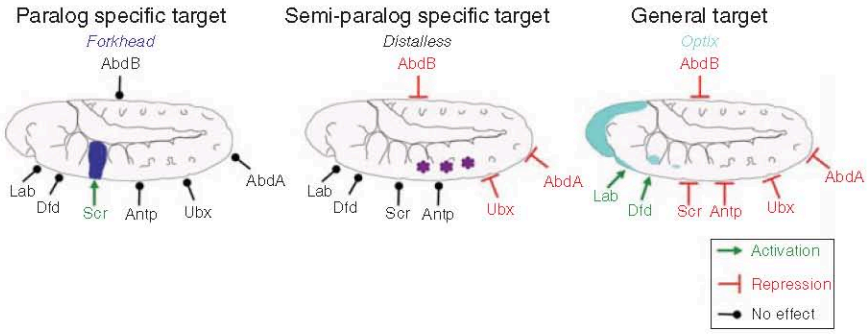


Figure 3.3 Three types of Hox target genes. "Paralog-specific" Hox target genes are those that are uniquely regulated by only a single Hox paralog, such as the activation of *fkh* by Scr. "Semi-paralog-specific" Hox target genes are those that are shared by a small subset of Hox paralogs, such as the repression of *Dll* by the abdominal Hox proteins Ubx, Abd-A, and Abd-B (schematized here is the *Dll304* embryonic enhancer element). "General" Hox target genes are those that are regulated by most, or perhaps all, Hox paralogs, such as the control of *optix* in *Drosophila*. Ideally, this classification should apply to individual *cis*-regulatory elements, not entire genes, to allow for the scenario that the same gene may fall into more than one of these categories (in two different tissues or times of development). For *fkh* and *Dll*, specific *cis*-regulatory elements that fit the "paralog-specific" and "semi-paralog-specific" criteria have been identified. In contrast, a single *cis*-regulatory element that is a "general" Hox target has not yet been identified and therefore remains hypothetical.

Author's personal copy

4. Hox Cofactors

Given that some Hox functions truly require a high degree of specificity, and that Hox homeodomains, themselves, are not sufficiently discriminating to account for this specificity, how is specificity achieved? One well-established way in which Hox proteins achieve specificity *in vivo* is to bind DNA cooperatively with other DNA-binding cofactors. To date, the best-characterized cofactors are all TALE (three amino acid loop extension) homeodomain proteins (Mann and Chan, 1996; Moens and Selleri, 2006). In *Drosophila*, the known TALE Hox cofactors are Extradenticle (Exd) and Homothorax (Hth). In the mouse, there are four Exd-related proteins (Pbx1, Pbx2, Pbx3, Pbx4) and five Hth-related proteins (Meis1, Meis2, Meis3, Prep1, and Prep2). In *Caenorhabditis elegans*, there are three genes encoding Exd-like proteins (*ceh-20*, *ceh-40*, *ceh-60*) and two that encode Hth-like proteins, *unc-62* and *psa-2*, which encodes for a truncated form that has no homeodomain (see Mukherjee and Bürglin, 2007 for a thorough description of the TALE family genes). Here, we collectively refer to Exd/Pbx/Ceh-20 as PBC proteins. These proteins all have the ability, at least on some DNA sequences, to bind with Hox proteins in a highly cooperative manner. In addition, it is important to stress that where it has been analyzed, TALE family homeodomain proteins also carry out many Hox-independent functions *in vivo* (Bessa *et al.*, 2008; Casares and Mann, 1998, 2001; Jiang *et al.*, 2008; Laurent *et al.*, 2007; Moens and Selleri, 2006). Because they have both Hox-dependent and Hox-independent functions the genetic analysis of TALE family genes needs to be interpreted with caution, since only a subset of the observed phenotypes is due to their role as Hox cofactors.

Protein interaction domains characterized in Hox, PBC, and Hth/Meis/Prep proteins have provided many insights into how these three factors assemble when bound to DNA. PBC proteins interact with Hth/Meis/Prep family members in a DNA-independent manner via highly conserved domains present in the N-terminal regions (PBC-A of PBC and HM of Hth/Meis) of these proteins (Mann and Affolter, 1998). In several cases, the nuclear localization and/or stability of these proteins has been shown to depend on this protein-protein interaction (Arata *et al.*, 2006; Berthelsen *et al.*, 1999; Haller *et al.*, 2004; Huang *et al.*, 2003; Mann and Abu-Shaar, 1996; Ryoo and Mann, 1999; Saleh *et al.*, 2000; Stevens and Mann, 2007). In contrast, PBC-Hox interactions appear to be more complicated—and potentially more interesting. The traditional view, which has been supported by biochemical, *in vivo*, and structural studies, is that a motif common to most Hox proteins—YPWM—makes direct contacts with the TALE motif in PBC homeodomains, which creates a hydrophobic pocket that binds the W in YPWM (Chan and Mann, 1996; Chang *et al.*, 1995; Joshi *et al.*, 2007; LaRonde-LeBlanc and Wolberger, 2003; Lu and Kamps, 1996;

Author's personal copy

Neuteboom *et al.*, 1995; Passner *et al.*, 1999; Phelan *et al.*, 1995; Piper *et al.*, 1999). For those Hox proteins that do not have an obvious YPWM motif, in particular the Abd-B paralogs, there is a conserved W residue that, at least for a subset of Abd-B paralogs, plays an important role in this protein-protein interaction (Shen *et al.*, 1997). However, more recent studies make it clear that there is more to PBC-Hox-DNA complex formation than the YPWM-TALE interaction. On the one hand, two studies have shown that mutation of the YPWM motif in Ubx fail to abolish cooperative binding with PBC proteins *in vitro* and some Ubx functions *in vivo* (Galant *et al.*, 2002; Merabet *et al.*, 2003; Shen *et al.*, 1997). On the other hand, a peptide immediately following the Ubx homeodomain, termed the UbdA motif because of its similarity between Ubx and Abd-A, is playing an important, and perhaps dominant, role in the interaction between these Hox proteins and Exd on some binding sites (Merabet *et al.*, 2007). Interestingly, consistent with the idea that the UbdA motif is playing a role in PBC-Ubx (and likely Abd-A) interactions, it contributes to Ubx functional specificity *in vivo* (Chan and Mann, 1993; Gebelein *et al.*, 2002). Although the UbdA motif is not found outside of arthropods, these findings suggest the more general idea that PBC proteins may have modes of interaction with other Hox proteins that are in addition to the classical YPWM-TALE interaction. The existence of multiple PBC interaction domains in a single Hox protein such as Ubx suggest that the way in which PBC-Hox complexes assemble onto *cis*-regulatory elements may vary from target to target, potentially expanding their ability to recruit additional transcription factors.

In addition to TALE family homeodomain proteins, the *Drosophila* homeodomain protein Engrailed (En) has also been shown to be a Hox cofactor (Gebelein *et al.*, 2004). In this case, En bound cooperatively with both Ubx and Abd-A to a regulatory element from the *Dll* gene, and En input is required for *Dll* repression in the posterior compartments of the abdominal segments (Gebelein and Mann, 2007; Gebelein *et al.*, 2004). Unlike the TALE cofactors, which can function with Hox proteins to both activate and repress target genes, it is likely that En-Hox dimers are more typically involved in gene repression due to En's ability to directly bind the corepressor Groucho (Jiménez *et al.*, 1997).

A subset of Zn finger-containing transcription factors, most prominently *Drosophila* Teashirt (Tsh), has also been suggested to be Hox cofactors (Robertson *et al.*, 2004; Taghli-Lamallem *et al.*, 2007). Although a very appealing idea, these factors do not seem to exhibit the same robust cooperativity that is typically observed between TALE factors and Hox proteins. And, at least in the one target where Tsh-binding sites were identified (*modulo*), they are not adjacent to the Hox-binding sites (Taghli-Lamallem *et al.*, 2007). Thus, at this time we prefer to classify these Zn finger factors as Hox "collaborators," which provide additional, essential inputs into a subset of Hox-targeted *cis*-regulatory elements (discussed in more detail below) (Table 3.1).

Table 3.1 Direct Hox-DNA-binding sites

Name (gene)	Hox	Reporter Validation	DNA Binding Site (Hox, PBC, Meis/Hth, Other)	Cooperative Cofactors	Other Factors	Target Origin; Notes	References
EphA2-repeat E (*EphA2/eck*)	Hoxa1 Hoxb1	*Ex vivo*	gcaTGATGGATGGgct	Pbx‡	ND	Mouse	(Chen and Ruley, 1998)
b1-ARE (*Hoxb1*)	Hoxb1 Lab	*In vivo*	ctcAGATGGATGGgctgcgggTGATTGAAGTgTCTTTGTCATGCTAAT gcttgggggTGATGGATGGgcgctggggCTGCCAaac	Pbx1, Prep1	Sox2, Oct1	Mouse; Autoregulatory element, R3 site (also known as Repeat3) functions as a Labial target in *Drosophila*	(Chan and Mann, 1996; Di Rocco et al., 2001; Ferretti et al., 2005; Pöpperl et al., 1995)
r4-Hoxa2 (*Hoxa2*)	Hoxb1	*In vivo*	ggaTGATTTATTTgag; attTGACAGtaatgaagagTGATAGATTGctc; ggcTGATGCATTAatt	Pbx1, Prep1	ND	Chicken and Mouse	(Tümpel et al., 2007)
b2-PP2 (*Hoxb2*)	Hoxb1	*In vivo*	gagCTGTCAgggggctaAGATTGATCGccc	Pbx1†, Prep1	ND	Mouse	(Ferretti et al., 2000)
A3-PP2; A3-PhP1 (*Hoxa3*)	Hoxb1	*In vivo*	ggtTGATTATTGacc; gagTCATAAATCTtgcccagccataaaTGACAAaaa	Pbx1a†, Prep1	ND	Mouse and Chicken; Mutation of Prep site did not affect reporter expression in vivo	(Manzanares et al., 2001)
P2b_0.38 (*Phox2b*)	Hoxb1	*In vivo*	gcgTGATTGAATtaa; ttaTTGTCAtgt	Pbx1a†, Prep1	ND	Mouse	(Samad et al., 2004)
FP8 (*COL5A2*)	Hoxb1	*Ex vivo*	gtcTGATTGATGGtaa	Pbx1a*	Prep1*	Human; Cotransfection of Prep1 increased expression ex vivo	(Penkov et al., 2000)
125bpHb9 (*Hb9*)	Hoxb1, Hoxb3	*In vivo*	agcTGATGAATTGacaaaaACTAATCA	Pbx1†	ND	Mouse	(Nakano et al., 2005)
EVIII (*CG11339*)	Lab	*In vivo*	tcgTGATCAATTAcagCTGACTggg	Exd, Hth	ND	*Drosophila*: Although the Exd half-site is required in vivo, cooperative binding of Lab/Exd dimers was not observed in vitro	(Ebner et al., 2005)
lab48/95 (*lab*)	Lab	*In vivo*	aatTGATGGATTGcccggcgccgaCTGTCAccg	Exd, Hth	ND	*Drosophila*: Autoregulatory element	(Ryoo et al., 1999)
enh450 (*ceh-13*)	CEH-13	*In vivo*	aaaTGATGGATGGttc	CEH-20	ND	*C. elegans*: Autoregulatory element	(Streit et al., 2002)
E2-A2RE (*Hoxa2*)	Hoxa2	*Ex vivo*	cggTGATTGATGGaag	Pbx1†	ND	Mouse, Autoregulatory element	(Lampe et al., 2004)
HRE1.25 (*Hoxa2*)	Hoxa2	*In vivo*	tctTGATTGATGAact	Pbx1a*	Prep1*	Mouse: Autoregulatory element, cotransfection of Prep1 increased expression ex vivo	(Lampe et al., 2008)
TTF-1-BS1&2 (*TTF-1*)	Hoxb3	*Ex vivo*	cctTAATTGgct; agaTAATTAgct	ND	ND	Rat; Both sites are necessary for expression ex vivo	(Guazzi et al., 1994)
CR3-HS1&2 (*Hoxb3, Hoxb4*)	Hoxb4, Hoxd4, Hoxb5	*In vivo*	agTGATTAATGGttttctgtaTAATTCtc	Exd†	ND	Mouse; Autoregulatory element; also responds to Dfd, Hoxb5, Scr, Antp, and Ubx in *Drosophila*	(Gould et al., 1997)
Rarb (*Rarb*)	Hoxb4, Hoxd4	*In vivo*	gggTGATAAATAAtgg	Pbx#	ND	Mouse; Role of Pbx not tested	(Serpente et al., 2005)
Repeat 3-TA (synthetic)	Dfd	*In vivo*	gggTGATTAATGGgcg	Exd	ND	Synthetic; (GG->TA) mutation of Repeat3 of b1-ARE	(Chan et al., 1997)
DfdEAE-module E (*Dfd*)	Dfd	*In vivo*	cccTAATTGccacgCATTAGctc	ND	Exd‡	*Drosophila*: Autoregulatory element, atypical Exd binding site, *exd* required in vivo	(Pinsonneault et al., 1997; Zeng et al., 1994)
rprDfd (*rpr*)	Dfd	*In vivo*	caaTAATTAccc; ctcTAATTGccc; aacTAATTGaca; tcaTAATTGagg	Not Exd‡	ND	*Drosophila*: Cooperative binding with Exd not observed	(Lohmann et al., 2002)
1.28DRE (*1.28*)	Dfd	*In vivo*	gttTAATTGgtt; tctTAATAGccg; ccaCTACATTAATTATgaa; ccgCGATAATAAatc	Not Exd‡	ND	*Drosophila*: Sites not tested individually, cooperative binding with Exd not observed	(Pederson et al., 2000)
jkl-216-site1 (*hlh-8*)	LIN-39	*In vivo*	agtTGAAaaATTACCgcg	CEH-20	ND	*C. elegans*	(Liu and Fire, 2000)
egl-17 (*egl-17*)	LIN-39	*In vivo*	ctcTGATTAATCActg	CEH-20	ND	*C. elegans*	(Cui and Han, 2003)
egl-18/elt-6-site 1 (*elg-18, elt-6*)	LIN-39	*In vivo*	gggTGATATATATgtt	CEH-20†	ND	*C. elegans*	(Koh et al., 2002)
Fkh250 (*fkh*)	Scr	*In vivo*	tcaAGATTAATCGcca	Exd, Hth	ND	*Drosophila*: Hth binding site not found	(Ryoo and Mann, 1999)
Lin-39-site2 (*lin-39*)	LIN-39	*In vivo*	catTGATTTATTTttg	CEH-20†	ND	*C. elegans*: Autoregulatory element	(Wagmaister et al., 2006)
mod-84 (*mod*)	Scr	*In vivo*	acaTAATTTgttagtatgTAATATtccctgGTATTTTTCCAATGactgtcaa gcagtgctgGTAAATGGCGGAACaag	ND	Tsh	*Drosophila*: Although both Scr and Tsh regulate *mod*, a cooperative interaction was not observed	(Taghli-Lamallem et al., 2007; Taghli-Lamallem et al., 2008)
p53 (*p53*)	Hoxa5	*Ex vivo*	tctTAATTCaaa	ND	ND	Human and Mouse	(Raman et al., 2000a)
PRP-62 (*PR*)	Hoxa5	*Ex vivo*	cggTAATTGggg	ND	ND	Human	(Raman et al., 2000b)
PTN-106 (*PTN*)	Hoxa5	*Ex vivo*	tttTAATAAgct	ND	ND	Human	(Chen et al., 2005)
HBE (*flk1*)	Hoxb5	*In vivo*	tgcTGATTAATGAaaa	Pbx#	ND	Mouse; Mutation affecting reporter activity located in putative Pbx site	(Wu et al., 2003)
bFGF (*bFGF*)	Hoxb7	*Ex vivo*	cgcTAATCTgg	ND	ND	Human	(Caré et al., 1996)
egl-1+5995 (*egl-1*)	MAB-5	*In vivo*	cgtTGATTTATTTtta	CEH-20	ND	*C. elegans*	(Liu et al., 2006)
apME680 (*ap*)	Antp	*In vivo*	ccaATTAtttTGATTAATGCcaa; cccATAAATAATATtaa; tttTTATGAgtt; ttgAAATGAact; tgaATTATGATTATTGCcat	Exd†	ND	*Drosophila*: *exd* required in vivo	(Capovilla et al., 2001)
Tsh (*tsh*)	Antp, Ubx	*In vivo*	aacTAATGTAATTACaaa; tgaTAATTGact; cacATAAATctt; tttTAATATttt	ND	ND	*Drosophila*: Cluster of Hox binding sites tested by deletion	(McCormick et al., 1995)
sal1.1 (*sal*)	Ubx	*In vivo*	ttaTAATgTGCCCGTCTTAATATgat	Not Exd or Hth	Mad/ Med	*Drosophila*: *exd* and *hth* not required in vivo	(Galant et al., 2002; Walsh and Carroll, 2007)
knot (*knot*)	Ubx	*In vivo*	gctTAATTTg; gctTAATTCt; cacTAATTAt; gccTAATTGTAATTGTTATTATTAATTAa	Not Exd or Hth	ND	*Drosophila*: Hox sites are additive, *exd* and *hth* not required in vivo	(Hersh and Carroll, 2005)
B3 (*βTub60D*)	Ubx	*In vivo*	ttcATAATTGcagcggccacactcCAATTAAatt	ND	ND	*Drosophila*	(Kremser et al., 1999)
CG13222-site1 (*CG13222*)	Ubx	*In vivo*	gtgTAATTTatc	Not Exd or Hth	ND	*Drosophila*: *exd* and *hth* not required in vivo	(Hersh et al., 2007)
Antp-P2 (*Antp*)	Ubx, AbdA	*In vivo*	30 Hox binding sites	ND	ND	*Drosophila*: Activity lost when all 30 sites are mutated.	(Appel and Sakonju, 1993)

(*Continued*)

Table 3.1 (*Continued*)

Name (gene)	Hox	Reporter Validation	DNA Binding Site (Hox, PBC, Meis/Hth, Other)	Cooperative Cofactors	Other Factors	Target Origin; Notes	References
DMX-R (*Dll*)	Ubx, AbdA	*In vivo*	gcaCTATAAAaCTGTCCgcggGAATGATTTAATTTcccaATATTgtc	Exd, Hth, En	Slp	*Drosophila*; Slp and En function as repressors	(Gebelein et al., 2004)
dpp-midgut enhancer (*dpp*)	Ubx, AbdA	*In vivo*	caaaTTTATTACTAATTGGGtgTGAATTGCaggcagtgcaagtgctgca tatgcagcatgcagcATAATCGAAatgggtgCTAATTGATag	Exd	ND	*Drosophila*; Atypical Exd binding site, additional Hox and Exd sites required	(Capovilla and Botas, 1998; Capovilla et al., 1994; Chan et al., 1994; Manak et al., 1994; Stultz et al., 2006; Sun et al., 1995)
RhoA (*rho*)	AbdA	*In vivo*	agtTCATTGATTGACATTTTTATTAtgc	Exd, Hth	Sens	*Drosophila*; Sens competes for binding with AbdA-Exd-Hth complexes	(Li-Kroeger et al., 2008)
XC-Box2 (*wg*)	AbdA	*In vivo*	gcaTAATCTAATTGcgg	Not Exd or Hth[a]	Mad, Medea, Creb	*Drosophila*; Exd and Hth do not form complexes with AbdA in vitro	(Grienenberger et al., 2003)
AH5 (*Hoxa4*)	Dfd, Scr, Antp, Ubx & AbdA	*In vivo*	ggCAATTAAATTTATGGgggcTATAATTActg	ND	ND	Mouse; Possible role for Exd, repressed by AbdA; activated by other Hox	(Haerry and Gehring, 1997)
FkhCON (synthetic)	Scr, Antp, Ubx, AbdA	*In vivo*	tcaAGATTTATGGcca	Exd, Hth	ND	Synthetic; Fkh250 binding site mutated to consensus Hox-Exd site, repressed by AbdA, activated by other Hox	(Ryoo and Mann, 1999)
MLC (*MLC1/3*)	Hoxc8, Hoxa10	*Ex vivo*	cttATTAAATTAcCATGTGtga	ND	ND	Rat; Factor binding to CATGTG (E-box) not identified, Hoxc8 activates and Hoxa10 represses	(Ceccarelli et al., 1999; Houghton and Rosenthal, 1999; Rao et al., 1996)
SBE (*OPN*)	Hoxc8, Hoxa9	*Ex vivo*	atgCAGTCtataaatgaaaagggtagtTAATGAcat	ND	Smad1, Smad3, Smad4	Mouse; Smad3 directly binds DNA site while Smad1 and Smad4 antagonize Hox binding	(Shi et al., 2001; Shi et al., 1999)
EphB4-1365 (*EphB4*)	Hoxa9	*Ex vivo*	agcTTATT	ND	ND	Human	(Bruhl et al., 2004)
E-510 (*E-selectin*)	Hoxa9	*Ex vivo*	atgCAATTTTATTAAtat	ND	ND	Human	(Bandyopadhyay et al., 2007)
N-CAM (*N-CAM*)	Hoxc6, Hoxb8, Hoxb9	*Ex vivo*	tTAATAATtac; cctTAATCAg	ND	ND	Mouse; Hoxc6 and Hoxb9 activate while Hoxb8 represses reporter expression ex vivo	(Edelman and Jones, 1995; Jones et al., 1993; Jones et al., 1992)
p21A10RE (*p21*)	Hoxa10	*Ex vivo*	tttTTATAAttt	Pbx1a*, Meis1b*	ND	Human; Cotransfection of Pbx1a and Meis1b increased reporter expression ex vivo, no direct binding sites identified	(Bromleigh and Freedman, 2000)
β3A (β_3-*integrin*)	Hoxa10	*Ex vivo*	aatGTTATTTtta	ND	ND	Human; Hox binding sites not confirmed by mutagenesis	(Daftary et al., 2002)
EMXC (*Emx2*)	Hoxa10	*Ex vivo*	tgtTTATGTgat	Not Pbx[a]	ND	Human; Pbx did not form complexes with Hoxa10 in vitro	(Troy et al., 2003)
Runx2-site1 (*Runx2*)	Hoxa10	*Ex vivo*	aggTTATAGctt	ND	ND	Rat	(Hassan et al., 2007)

Name (gene)	Hox	Reporter Validation	DNA Binding Site (Hox, PBC, Meis/Hth, Other)	Cooperative Cofactors	Other Factors	Target Origin; Notes	References
PPE (*Ren-l*[c])	Hoxd10	*In vivo*	tgtTTCCACAct; cgcTTCCggc; cttTGATTTATTAccc	Pbx1b, Prep1*	N1IC, Ets-1	Mouse; Homeodomain of Prep1 not necessary for activation ex vivo, Prep1 binding site not identified	(Glenn et al., 2008; Pan et al., 2005; Pan et al., 2001)
Six2 (*Six2*)	Hox11	*In vivo*	gTTATCTgacccggggcctgcccgcgccaGacaatagTCGaGTCA aattaTTc	ND	Pax2, Eya1	Mouse	(Gong et al., 2007)
Enpp2 (*Enpp2*)	Hoxa13	*Ex vivo*	TTAATTG; TTAACAT; TTTATAT	ND	ND	Mouse	(McCabe and Innis, 2005)
Sostdc1 (*Sostdc1*)	Hoxa13	*Ex vivo*	15 sites tested	ND	ND	Mouse	(Knosp et al., 2007)
EphA7-site 3 (*EphA7*)	Hoxa13 Hoxd13	*Ex vivo*	ataTTATTGgag	ND	ND	Mouse	(Salsi and Zappavigna, 2006)
hHa2-motifs: 1, 3, 6, 11 (*hHa2*)	Hoxc13	*Ex vivo*	aaaTTAATTAgcag; tctTTTATTGgaa; cttTTAATGAaag; catTTAATATgttg	Not Meis/Prep	ND	Human; Meis/Prep localization cytoplasmic	(Jave-Suárez and Schweizer, 2006; Jave-Suárez et al., 2002)
hHa5-motifs: 2, 3, 4, 6, 9, and 13 (*hHa5*)	Hoxc13	*Ex vivo*	actTTAATGAgga; gttTTAATAGaaa; aggTTAATGAggg; tatTTTATGAgact; tctTTTATTGgcct; cagTTAATTGgac	Not Meis/Prep	ND	Human; Meis/Prep localization cytoplasmic	
hHa7-motif 2 (*hHa7*)	Hoxc13	*Ex vivo*	gctTTAATGAgct	ND	ND	Human	
Foxq1 (*Foxq1*)	Hoxc13	*Ex vivo*	accTTCATTAcaa; tccTCCATAAAaca; agcTTAATAGGgac	ND	ND	Mouse; Other Hoxc13 sites may be required	(Potter et al., 2006)
BE3-6 (*yellow*)	AbdB	*In vivo*	aggTCGTAAAAcgtattttacccatttgcatgtTTATTATGcgt	ND	ND	*Drosophila*	(Jeong et al., 2006)
BabDE (*bab*)	AbdB	*In vivo*	7 TTTAT, 7 TTTAC, 8 TTAT, and 2 ACAATGT binding sites	ND	Dsx	*Drosophila*	(Williams et al., 2008)
i4c-S2 (*psa-3*)	NOB-1	*In vivo*	tttTGATAGTAATttt	CEH-20	ND	*C. elegans*; CEH-20 facilitates NOB-1 DNA binding independent of its own binding site	(Arata et al., 2006)

[a] Tested only *in vitro*.
[b] Half-site not directly tested.
[c] Tested *ex vivo*.
[d] PBC-binding site recognized in sequence but untested.

Targets are ordered according to Hox paralog and are derived from human, mouse, chicken, *Drosophila*, and *C. elegans*. All of these elements have been shown to require direct Hox input for activity either *in vivo* or *ex vivo* (cell culture). Colored shading in the first column indicates if the element is repressed, activated or both (red, green, and blue, respectively) by Hox input. Nonadjacent binding sites within the same enhancer are separated by semicolons. Underlined sites are shared between two proteins using the colors defined in the DNA-binding site column. ND, not determined.

Author's personal copy

Given the high degree of specificity required for some Hox functions, and that there are dozens of *Hox* genes in vertebrates, it is perhaps surprising how few bona fide Hox cofactors have been identified. One answer to this paradox, discussed later in this chapter, comes from a recent atomic-level resolution view of how TALE cofactors bind to DNA with Hox proteins. Another possible answer, especially in vertebrates, is that multiple *Pbx* and *Meis/Prep* genes encode for proteins with different biochemical properties and thus expand the number of Hox cofactors. Perhaps analogously, although *Drosophila* has a single *exd*-like gene and a single *hth*-like gene, alternative splicing of *Drosophila hth* also adds to the repertoire of Hox cofactors present in the fly (Noro *et al.*, 2006). Specifically, *hth* encodes both homeodomain-containing and homeodomain-less (HDless) isoforms. Not only do both of these isoforms contribute to *hth* functions, there is a clear division of labor for these isoforms. For example, the homeodomain-containing form of Hth is not required for a large number of embryonic functions, including many (but not all) Hox-dependent functions. In contrast, the homeodomain-containing form is essential for *hth* to specify antennal development, which is one of its Hox-independent functions. The fact that these two Hth isoforms exist suggests the possibility that they may also be used in different ways to achieve Hox specificity. Strikingly, in *C. elegans*, the gene *psa-3* encodes an HDless Hth/Meis ortholog (Arata *et al.*, 2006). Thus, *C. elegans* produces a very similar HDless isoform, but via gene duplication and truncation, instead of by alternative splicing as in *Drosophila* and vertebrates. The presence of HDless isoforms of Hth/Meis in *C. elegans*, *Drosophila*, and vertebrates reinforces the idea that it carries out critical functions that are distinct from those executed by homeodomain-containing isoforms.

Although *exd* appears to produce only a single isoform, some of the vertebrate *Pbx* genes produce multiple isoforms via alternative splicing (Milech *et al.*, 2001; Monica *et al.*, 1991; Wagner *et al.*, 2001). Using the yeast two-hybrid assay, some evidence exists that a subset of isoforms have distinct abilities to interact with Meis 1, Meis 2a, and Prep1 (Milech *et al.*, 2001). Such differences may also be important for Hox specificity, and may be reflected in the arrangement of binding sites for Hox and TALE proteins in Hox-targeted *cis*-regulatory elements.

5. What Do *In Vivo* Hox-Binding Sites Look Like?

An important approach to understand how Hox proteins regulate target gene expression, and to reveal potential generalizations, is to examine the *cis*-regulatory elements they directly bind to *in vivo*. Once a set of *in vivo*-validated Hox-targeted *cis*-regulatory elements are in hand, several

Author's personal copy

questions can be asked. These include: How many also require input from known cofactors? How many Hox-binding sites are present in each element?, and What other regulatory inputs are there? To provide initial answers to these questions, we have surveyed the literature with the goal of cataloging the majority of the direct Hox-binding sites that have been examined to date, in both vertebrates and invertebrates. By "direct," we included in this survey only Hox-binding sites that have been shown by a reporter gene assay (in cell culture (*ex vivo*) or *in vivo*) to be required for the activity of a *cis*-regulatory element (Table 3.1). Therefore, some recent genome-wide studies fell short of these stringent criteria for validation (Ebner *et al.*, 2005; Hueber and Lohmann, 2008; Hueber *et al.*, 2007; McCabe and Innis, 2005). Below, we discuss the results of this survey and their implications for Hox specificity.

We found 66 *cis*-regulatory elements for which there is strong experimental evidence for direct and essential Hox input (Table 3.1). Of these, 29 have been shown to use PBC cofactors (Exd/Pbx/Ceh-20). Two additional elements appear, by sequence, to have PBC-Hox composite sites, making a total of 31 elements with PBC-Hox sites in this data set (Table 3.1). For seven elements, there is experimental evidence that they do not use these cofactors. The remaining 30 targets have not been directly tested for PBC input, although two of these have been shown not to use Hth/Meis or Prep proteins (Table 3.1). Finally, among the 66 targets in this list, there are 11 examples in which other direct inputs have been shown to be required for the activity of the *cis*-regulatory element.

Although we need to cautiously interpret this relatively small number of elements, several interesting features emerge by analyzing these examples. First, it is noteworthy that for a large fraction of the elements (29 of the 36 elements where it was examined) Hox proteins bind their binding sites cooperatively with a PBC factor (Table 3.1). This is likely an overestimate of the frequency of PBC-Hox-binding sites, because sequence gazing of the 30 elements that were untested for PBC input suggests that many do not have an obvious PBC-binding site. Nevertheless, the abundance of PBC-Hox composite-binding sites in this list underscores the widespread contribution of these cofactors to Hox-binding *in vivo*. Strikingly, with only two exceptions (*Dll* DMX and *fkh250*^{con}), PBC-Hox-binding sites are used for gene activation, not repression. In contrast, the Hox sites that do not have clear PBC input are used for both repression and activation. If this overall correlation continues to hold up, it suggests that PBC-Hox complexes are, in general, more likely to recruit transcriptional coactivators rather than corepressors.

Second, there is a trend for the anterior Hox proteins (paralogs 1-5) to use PBC cofactors more than the posterior Hox proteins (paralogs 6-13) (Table 3.1). Of the 30 elements targeted by a Hox 1-5 paralog, 20 have a required PBC-Hox-binding site. In contrast, of the 36 elements targeted

Author's personal copy

by a Hox 6-13 paralog, only nine have been shown to have an essential PBC-Hox-binding site.

Third, elements that do not have PBC input are more likely to have multiple Hox-binding sites than elements that have PBC input (Table 3.1). For those elements that use Hox-PBC sites, the average number of Hox-binding sites is 1.2 (ranging from 1 to 3), whereas for those elements that do not appear to use Hox-PBC sites, the average number of sites is 2.8 (ranging from 1 to 30). Thus, from this data set, it appears that Hox sites without PBC input often function in groups. If this trend holds up, it may reflect a lower affinity for non-PBC Hox sites when compared to PBC-Hox sites. Perhaps multiple non-PBC Hox sites are therefore required in an additive manner to elicit a transcriptional response.

Although many of the targets listed in Table 3.1 have Hox-binding sites that do not appear to have direct PBC input, we avoid calling them Hox "monomer" sites because it is plausible that currently unidentified factors bind with Hox proteins (cooperatively or noncooperatively) to these sites. In fact, when true Hox "monomer"-binding sites—synthetic, but high-affinity-binding sites—were used to drive a reporter gene in *Drosophila* embryos, they did not produce expression patterns consistent with their ability to bind dozens of homeodomain proteins (Vincent *et al.*, 1990). This experiment, together with the analysis of *in vivo* Hox targets listed in Table 3.1, suggest that Hox proteins never work as monomers.

The arrangement of binding sites also varies in interesting ways within this data set. Of the 40 PBC-Hox sites (distributed among 31 elements), 33 have the PBC half-site adjacent to the Hox half-site, and the majority of these (26) have the structure nnATnnATnn (where the first and second ATs form the core of the PBC and Hox half-sites, respectively). In one case (DMX-R from *Dll*) the PBC-Hox site has the structure nnATnnnATnn and there is one example (in the *dpp* midgut element) where the Hox site precedes the PBC site (Table 3.1). It is possible that these atypical arrangements help these sites be more selective for a subset of Hox proteins. It is also possible that the unique three-dimensional architectures of the protein complexes assembled by these atypical PBC-Hox-binding sites are important for recruiting additional, element-specific transcriptional effectors. Consistent with these ideas, both *in vitro* Hox-binding specificity and *in vivo* activity of DMX-R were reduced when the spacing between the two half-sites was changed from 3 to 2 bp (Gebelein *et al.*, 2002). Of the 40 PBC-Hox sites, eight have been shown to have a nearby Hth/Meis or Prep-binding site. Although the low number of identified Hth/Meis/Prep-binding sites may in part be because they are not always looked for, it may also reflect the fact that there are isoforms of Hth and Meis that do not have a homeodomain and thus would not be expected to make DNA contacts.

Because both PBC-Hox and Hth/Meis/Prep-binding sites have a clear orientation, four possible arrangements of these two binding sites are

Author's personal copy

possible. Interestingly, of these four, only one is not observed (TGA-CAG. . .PBC-Hox, where TGACAG represents one orientation of a Hth/Meis site) (Table 3.1). That three of the potential orientations have been observed suggests an inherent flexibility in how these complexes can bind DNA. Further, because the different orientations are expected to orchestrate the assembly of protein-DNA complexes that have unique three-dimensional architectures, these observations also suggest the possibility that they have unique biochemical properties, such as their ability to recruit additional transcriptional coactivators or corepressors.

An unusual arrangement of binding sites is found in an Abd-A-targeted element from the *Drosophila rhomboid* (*rho*) gene (Table 3.1) (Li-Kroeger *et al.*, 2008). In this element, the order of binding sites is PBC-Hth-Hox. Robust cooperative DNA binding to this element was observed between a preformed Exd-Hth dimer and Abd-A. In principle, because both Exd and Hth are TALE homeodomain proteins, they both have the ability to bind the "YPWM" motif of Abd-A, raising the question of which, if either, TALE motif Abd-A is interacting with. Interestingly, mutation of the Hth site, but not the Exd site, dramatically reduced complex formation (Li-Kroeger *et al.*, 2008), suggesting that the Hth-binding site and, perhaps, the Hth TALE motif are more critical for complex formation. Perhaps analogously, a direct Hox-Hth interaction was also proposed to exist in the repressor element from the *Dll* gene (Table 3.1) (Gebelein *et al.*, 2004). The existence of these atypical arrangements suggests that there may be additional flexibility in how Hox, PBC, and Meis/Hth/Prep proteins assemble onto target DNAs. The dissection of the *rho* element in particular emphasizes that carrying out careful mutagenesis and follow-up *in vivo* studies will be critical for identifying additional novel architectures that are used by these factors *in vivo*.

As discussed above, two recent reports described the *in vitro* binding site preferences for nearly all mouse and fly homeodomains, including all Hox homeodomains (Berger *et al.*, 2008; Noyes *et al.*, 2008). How do these results compare with the *in vivo* Hox-binding sites listed in Table 3.1? To answer this question, we generated binding site logo diagrams using the B1H-derived binding sites for the *Drosophila* Hox homeodomains (Noyes *et al.*, 2008) and the *in vivo* Hox-binding sites listed in Table 3.1 (Fig. 3.2). The B1H-derived logos are all based on at least 19 individual binding sites, while the number of individual *in vivo* binding sites for each Hox protein that went into this analysis ranged from 12 (for Scr/Hox5) to 57 (for Ubx/Hox7). Side-by-side analysis of these two sets of sequences reveals some noteworthy differences (Fig. 3.2). First, consistent with the high proportion of PBC-Hox sites in the Hox1/Labial targets, the *in vivo* consensus sequence readily identified a PBC half-site (TGAT). In addition, while the B1H selection tended to identify TAATTA for Hox1/Labial, GGATGG is commonly observed in the *in vivo* data set for these Hox proteins. Other,

Author's personal copy

though less dramatic, differences are also observed between the B1H and *in vivo* data sets for nearly all of the Hox paralogs (Fig. 3.2). These comparisons reinforce the view that the *in vivo* environment, due to the presence of other cofactors, collaborators, or differences in DNA and/or chromatin structure, influences Hox-binding site preferences.

6. Insights into Hox Specificity from Structural Studies

Several monomeric homeodomain-DNA structures have been solved, and all reveal a very similar mode of DNA recognition by this DNA-binding domain (Gehring *et al.*, 1994). Briefly, the third alpha-helix, also called the recognition helix, lies in the major groove of the DNA, where it makes several direct and water-mediated contacts with specific bases and the phosphate backbone. Ile47, Gln50, Asn51, and Met54, residues that are present in all Hox homeodomains, are primarily responsible for making these contacts. In addition, the so-called N-terminal arm, which precedes the first alpha-helix, is typically observed in the minor groove. Arg5, an N-terminal arm residue present in nearly all homeodomains, is the most commonly observed residue in the minor groove.

In addition to these monomeric homeodomain-DNA structures, we now have X-ray structures of five different PBC-Hox-DNA complexes (Fig. 3.4). The Hox homeodomains in these structures recognize the DNA using the same contacts that were observed in the monomeric structures, demonstrating that the presence of PBC does not grossly alter the way in which Hox homeodomains bind DNA. In all five of the PBC-Hox structures (with four different Hox proteins: Hoxb1, Scr, Ubx, and Hoxa9), the PBC and Hox homeodomains bind DNA in a head-to-tail orientation, with very similar overall arrangements. In all five structures the Hox YPWM motif binds the hydrophobic TALE pocket in the PBC homeodomain. Thus, these four Hox proteins have the capacity to bind DNA cooperatively with PBC proteins using very similar protein-protein and protein-DNA contacts. However, we note that the currently available structures provide an incomplete picture because, for some binding sites, other protein motifs, such as the UbdA motif of Ubx and Abd-A, play an important role in forming PBC-Hox-DNA complexes (Merabet *et al.*, 2007). Currently, no structural information exists about these domains or how they contact PBC proteins.

PBC proteins not only bind cooperatively to DNA with Hox proteins, they also increase Hox-DNA-binding selectivity. This phenomenon is best illustrated with a few examples. In the absence of cofactors, the Hox1/Labial paralog shows a preference for binding the sequence TAATTA

Author's personal copy

(Fig. 3.2) (Berger *et al.*, 2008; Noyes *et al.*, 2008). In the presence of a PBC protein, a PBC-Hox1/Lab heterodimer prefers to bind the sequence TGAT[t/g]GATgg, where [t/g]GATgg is the Hox-binding site (base pairs with brackets indicate multiple possibilities at the same position and lowercase letters indicate only partially preferred base pairs) (Fig. 3.2). In contrast, while Ubx also prefers to bind TAATTA as a monomer, it will readily bind TGATTTATTT as a PBC-Ubx heterodimer, where TTATTT is the Hox-binding site (Berger *et al.*, 2008; Noyes *et al.*, 2008) (Fig. 3.2). Thus, the presence of PBC changes the DNA-binding preferences of both Labial/Hox1 and Ubx, but toward different sequences.

These observations raise the question of how the same cofactor can generate two different outcomes for these two Hox paralogs. The likely answer is that the specificity information is in the Hox protein, but is only revealed in the presence of the cofactor. Two recent crystal structures of PBC-Hox-DNA complexes support this idea (Joshi *et al.*, 2007). In one, an Exd-Scr heterodimer is bound to a paralog-specific PBC-Hox-binding site (*fkh250*) from the *fkh* gene, an *in vivo* Scr target gene. A second crystal structure shows the same two proteins bound to a consensus PBC-Hox sequence (*fkh250*^{con}). Importantly, additional protein-DNA contacts were observed in the Exd-Scr-*fkh250* complex, but not in the Exd-Scr-*fkh250*^{con} complex. These contacts are derived from the N-terminal arm of the Scr homeodomain and, surprisingly, a nonhomeodomain residue in the linker region between Scr's YPWM motif and homeodomain. Both of these side chains are inserted into the minor groove of the *fkh250*-binding site (Figs. 3.4 and 3.5) (Joshi *et al.*, 2007). Two additional observations are of interest. First, the DNA minor groove where these two side chains insert is unusually narrow, and significantly narrower than the analogous region of the *fkh250*^{con}-binding site (Fig. 3.5). This suggests that subtle differences in DNA structure are likely to contribute to Hox-binding specificity. Second, the residues making these minor groove contacts, which are part of a normally unstructured region of Scr, require the YPWM-Exd interaction to be positioned in close proximity to the minor groove. Thus, the paralog-specific binding of Scr to its binding site in *fkh* depends on three contributing features (1) an unusual DNA structure; (2) paralog-specific residues in the Scr homeodomain and linker that insert into this DNA structure; and (3) the YPWM-Exd interaction, which positions the N-terminal arm and linker region so that this normally unstructured peptide can make these contacts.

Although these structures provide insights into why Exd-Scr specifically binds *fkh250*, they raise the question of how general these findings are. Two additional observations suggest that the underlying principles revealed by these structures may, in fact, be a general feature of PBC-Hox-DNA interactions, at least for paralog-specific and semi-paralog-specific target sites. For one, Hox N-terminal arms and linker regions are evolutionarily conserved in a paralog-specific manner (Joshi *et al.*, 2007; Mann, 1995;

Author's personal copy

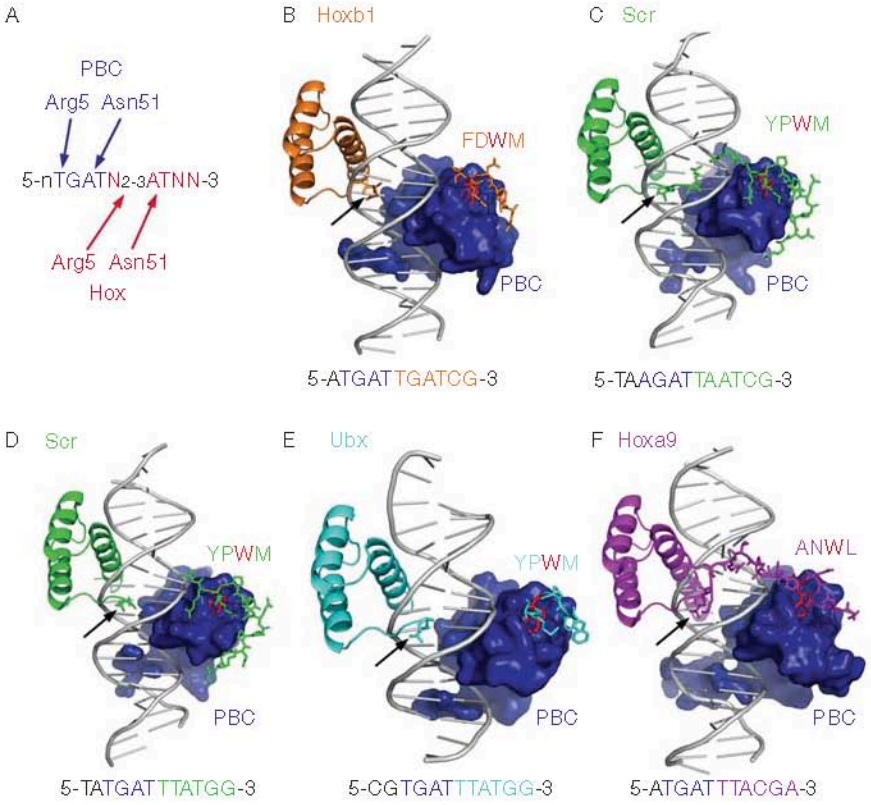


Figure 3.4 Common and unique features of PBC-Hox-DNA complexes. (A) Consensus PBC-Hox-binding sites have a PBC half-site (typically TGAT or AGAT, blue) and a Hox half-site (typically NNATNN, red). Minor groove (Arg5) and major groove (Asn51) contacts observed in all five of the PBC-Hox crystal structures are indicated. N_{2-3} reflects the observation that the PBC and Hox Asn51-contacted "AT" are usually separated by 2 bp, but 3 bp spacings have also been observed. (B)-(F) Overviews of the five existing PBC-Hox-DNA crystal structures. In all examples, the PBC protein (for most examples, just its homeodomain) is shown as a blue surface. The Hox proteins, which include the YPWM motif (which is FDWM in Hoxb1 and ANWL in Hoxa9), linker, and homeodomain, are color-coded as indicated. Only side chains around the YPWM, linker, and N-terminal arm are shown; homeodomain helices and loops are shown in cartoon format. The Trp (W) in the YPWM motif is colored red in all cases to indicate its conserved interaction with the TALE motif in the PBC homeodomain. In all cases, Arg5 of the Hox N-terminal arm is observed in the minor groove (black arrows). In only two cases (C; Exd-Scr bound to *fkh250* and F; Pbx-Hoxa9 bound to a consensus sequence) are additional N-terminal arm and linker regions observed; these regions are disordered in the other three structures. The DNA sequences present in these structures are shown below the structure, with the PBC and Hox half-sites color-coded. These images were generated using PyMol; the PDB accession numbers for these structures are (B) 1B72, (C) 2R5Z, (D) 2R5Y, (E) 1B8I, and (F) 1PUF.

Author's personal copy

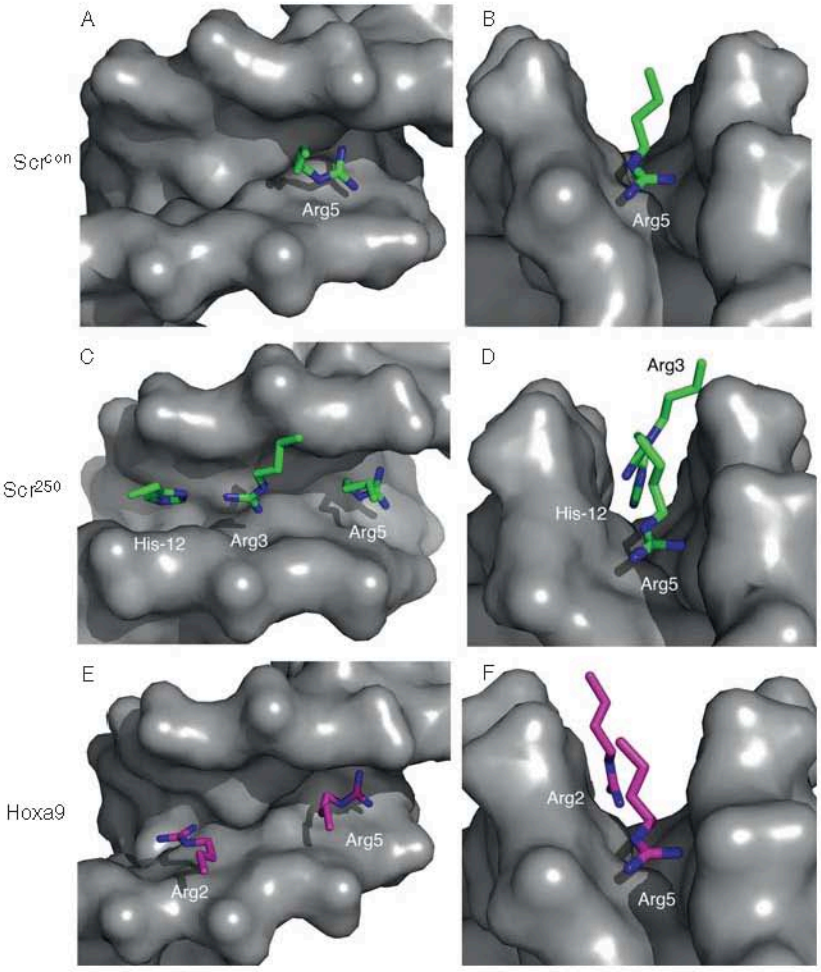


Figure 3.5 Interactions between Hox proteins and the DNA minor groove. Shown are images from X-ray crystal structures of PBC-Hox-DNA complexes, focused only on the interaction between the minor groove (shown as the gray surfaces) and the amino acid side chains of N-terminal arm/linker residues. The left-hand images (A, C, E) look into the minor groove from the top; the right-hand images (B, D, F) look along the axis of the minor groove. (A, B) Exd-Scr bound to the *fkh250*^{con} consensus-binding site. Only Arg5 from the N-terminal is observed in the minor groove. (C, D) Exd-Scr bound to the *fkh250 in vivo* binding site. In contrast to the *fkh250*^{con} structure, Arg3 (from the N-terminal arm) and His-12 (from the linker) are observed in the minor groove, in addition to Arg5. Note also that the minor groove in the *fkh250* structure appears narrower than in the *fkh250*^{con} structure (compare B with D). See Joshi *et al.* (2007) for details. (E, F). In the Pbx-Hoxa9 structure, one additional N-terminal arm residue, Arg2, is observed, together with Arg5. The Hoxa9 linker is unusually short (four residues), and none of them are seen inserting into the minor groove. See LeRonde-LeBlanc and Wolberger (2003) for details.

Author's personal copy

Morgan *et al.*, 2000). The two minor groove inserting residues in Scr are conserved in all Hox5 paralogs, while other Hox paralogs have different N-terminal arm and linker sequences that are also equally conserved in a paralog-specific manner. Second, another PBC-Hox structure, that of Pbx-Hoxa9, also has significant N-terminal arm-minor groove contacts (LaRonde-LeBlanc and Wolberger, 2003) (Figs. 3.4 and 3.5). Although the binding site used in the Pbx-HoxA9 structure is not from an *in vivo* HoxA9 target, the N-terminal arm-minor groove contacts are also dependent on the YPWM-Pbx interaction, illustrating the potential generality of DNA minor groove recognition by PBC-Hox complexes. The Pbx-Hoxa9 structure also reveals significantly more contacts with the phosphate backbone of the DNA, perhaps accounting for its higher affinity compared to other PBC-Hox complexes (LaRonde-LeBlanc and Wolberger, 2003).

Along the same lines, it is also noteworthy that linker length—and, consequently, the distance between the YPWM motif and the N-terminal arm—varies significantly among Hox proteins. Not only are there huge differences in linker lengths (ranging from >50 in Labial to <5 in Hoxa9), linker length roughly correlates with Hox paralog: anterior (3′) Hox paralogs have a much greater tendency for long linkers than more posterior (5′) Hox paralogs. In addition, Hox linkers also vary for individual paralogs due to alternative splicing; the Ubx linker, for example, ranges from 8 to 51 depending on the Ubx isoform (Kornfeld *et al.*, 1989; O'Connor *et al.*, 1988). Consequently, the distance between the YPWM motif and the homeodomain varies and would therefore be expected to affect how the N-terminal arm and/or linker region interacts with the DNA. These intriguing observations contribute to the idea that DNA contacts made by linker and N-terminal arm residues may be generally critical for paralog-specific and cofactor-dependent DNA binding. These observations are also consistent with the alternative idea that linker residues interact with additional proteins, although such factors have not yet been identified (Merabet *et al.*, 2003).

In summary, the common and unique features revealed in these five PBC-Hox-DNA structures, together with previously solved monomeric homeodomain-DNA structures, suggest that DNA recognition by Hox proteins uses two tiers of information that provide different degrees of specificity (Fig. 3.6). The first tier uses DNA-contacting residues that are common to all Hox proteins (Arg5, Ile47, Gln50, Asn51, and Met54) to promote Hox binding to "AT"-rich sequences, such as TAAT[gt][ga]. The second tier uses additional DNA-contacting side chains that come from the N-terminal arm and linker regions; these contacts are cofactor-dependent and paralog-specific (Fig. 3.6). Further, at least for the case of Exd-Scr bound to *fkh250*, these side chains recognize a DNA structure, rather than a specific DNA sequence. The fact that N-terminal arm and linker residues are conserved in a paralog-specific manner suggests that the two

Author's personal copy

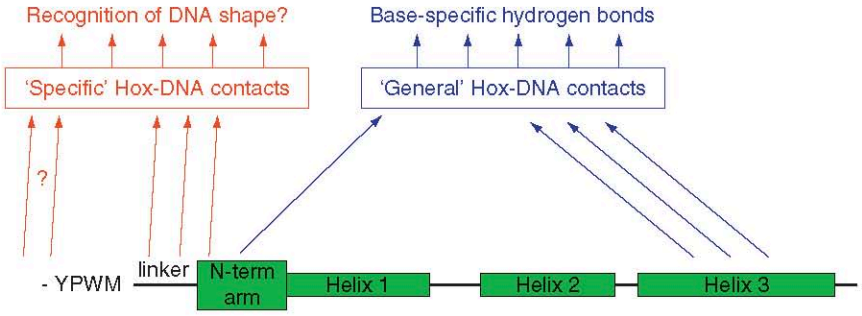


Figure 3.6 Two tiers of Hox-DNA-binding specificity. Hox proteins bind DNA using two levels of protein-DNA contacts. DNA contacts made by Arg5 (in the N-terminal arm) and Ile47, Gln50, Asn51, Met54 (in the third helix) are used by all Hox proteins to bind "AT"-rich DNA sequences ("general" Hox-DNA contacts), but are not good at distinguishing between Hox paralogs. With the help of cofactors (such as PBC proteins), paralog-specific DNA contacts are mediated by linker and N-terminal arm residues. "General" DNA contacts make hydrogen bonds in the DNA major groove. "Paralog-specific" DNA contacts may read a DNA structure, such as the narrow minor groove seen in the Exd-Scr-*fkh250* structure.

tiers of recognition may underlie the recognition of paralog-specific DNA sequences by other Hox-cofactor complexes.

7. Activity Regulation of Hox Proteins: The Role of Hox Collaborators

Although TALE family proteins clearly play an important role in DNA-binding site recognition, Hox proteins use these cofactors to both activate and repress target genes, raising the question of how gene activation versus repression is determined. Although there is currently only one example, one answer is that Hox proteins may use dedicated repressors, such as En, as Hox cofactors in gene repression (Gebelein *et al.*, 2004). Another possibility, which will be no surprise to people used to thinking about *cis*-regulatory elements, is that additional factors bind to Hox-targeted elements and contribute to their activities. Given the increasing number of directly regulated Hox targets that have been characterized, several such accessory factors, which we refer to here as Hox collaborators, have been identified. We have classified a factor as a Hox collaborator if it provides a direct and essential input into a Hox-regulated element, but has not been definitively shown to bind DNA cooperatively with Hox proteins (Table 3.1). These factors include the *Drosophila* Forkhead domain protein Sloppy paired (Slp), which collaborates with Ubx and Abd-A to repress *Dll* in the *Drosophila* abdomen (Gebelein *et al.*, 2004).

Author's personal copy

Interestingly, recent results suggest that a vertebrate ortholog of Slp, FoxP1, collaborates with Hox proteins during the establishment of motor neuron identities in the mouse (see Chapter 1) (Dasen *et al.*, 2008; Rousso *et al.*, 2008). Thus, the collaboration of FoxP1/Slp (and perhaps other Forkhead domain factors) with Hox proteins appears to be evolutionarily conserved. Like En, Slp may be a dedicated transcriptional repressor due to its Groucho interaction domain, thus providing a mechanism to explain the regulatory sign of *Hox-Slp*-targeted genes. In addition to these cases, protein-protein interaction and genetic studies suggest that the range of potential Hox collaborators is extensive (Kataoka *et al.*, 2001; Luo *et al.*, 2004; Plaza *et al.*, 2008; Prévôt *et al.*, 2000).

Transcription factors that provide cells with identity information, like the Hox factors, have been generally referred to as Selector transcription factors, a term originally coined 40 years ago to describe common properties of the genes *Ubx* and *en* (García-Bellido, 1975; Mann and Carroll, 2002; Mann and Morata, 2000). More recently, selector proteins have been proposed to frequently, if not always, function together with effector transcription factors that are downstream of cell-cell signaling pathways (reviewed previously by Bondos and Tan, 2001; Curtiss *et al.*, 2002; Mann and Affolter, 1998). Thus, not surprisingly, as more elements that are directly targeted by Hox proteins are dissected, signal effector transcription factors are being identified as Hox collaborators. In particular, vertebrate and *Drosophila* SMADs, effectors of the TGF-beta and Decapentaplegic (Dpp) pathways, have been identified as Hox collaborators in several *cis*-regulatory elements (Galant *et al.*, 2002; Grienenberger *et al.*, 2003; Shi *et al.*, 1999, 2001; Walsh and Carroll, 2007) (Table 3.1). Although SMAD-Hox-DNA-binding cooperativity has not been described, there are several reports suggesting that SMADs and Hox proteins can directly interact with each other (Wang *et al.*, 2006; Williams *et al.*, 2005; Zhou *et al.*, 2008). Such interactions may be critical for building an enhanceosome-like structure on Hox-targeted *cis*-regulatory elements. Although the number of examples shown to have direct inputs by signaling effectors is currently low, genetic analyses suggest that this phenomenon is likely to be a general feature of Hox-targeted *cis*-regulatory elements, and will probably extend to other signaling pathways, including Hedgehog (Hh), Wnts, and Notch (Arata *et al.*, 2006; Crickmore and Mann, 2007, 2008; Hersh *et al.*, 2007; Joulia *et al.*, 2006; Marty *et al.*, 2001; Merabet *et al.*, 2005; Weatherbee and Carroll, 1999).

Like Hox cofactors, the presence of a particular Hox collaborator does not guarantee the sign of the transcriptional regulation. In the *sal1.1* element, for example, Ubx collaborates with Mad/Medea to repress transcription, while in the *XC* midgut element from the *ug* gene, Mad/Medea collaborates with Abd-A to activate transcription (Grienenberger *et al.*, 2003; Walsh and Carroll, 2007). This difference is not simply due to different Hox paralogs, because both Ubx and Abd-A can both directly

Author's personal copy

repress and directly activate transcription (Table 3.1). Instead, these observations imply that additional, currently unknown, factors are being recruited to these elements to determine the sign of the transcriptional regulation.

Based on these direct examples, together with the larger number of genetically defined examples of Hox—signaling collaborations, we suggest that it may be a general feature of the multiprotein complexes that are built on Hox-targeted *cis*-regulatory elements. In fact, because Hox proteins work in so many different developmental contexts, it is likely that Hox collaborators will ultimately include a very large number of different types of transcription factors. Perhaps the ability of Hox proteins and PBC-Hox dimers to interact with a large number of different collaborators makes these proteins such ideal regulators of cell type and tissue identities.

Yet, despite this flexibility, it is critical to stress that the Hox factors play the central role in the function (and/or the assembly) of these multiprotein complexes, because without them, these complexes cannot function. Moreover, for paralog-specific functions, the activity and/or assembly of these complexes must depend on the correct Hox paralog and cofactors. Because of their central role, we would therefore like to coin the term "Hoxasome" to describe these multiprotein complexes, which include the Hox proteins, their cofactors, and their collaborators.

8. Insights into Hoxasome Function from *cis*-Regulatory Element Architecture

One straightforward view for how Hoxasomes function is that, once assembled, they recruit coactivators, corepressors, and/or chromatin remodeling complexes that ultimately carry out transcriptional regulation much like any other enhanceosome. Indeed, consistent with this view, there have been numerous reports describing direct interactions between Hox proteins and/or TALE cofactors with these more general components of the transcriptional machinery (Chariot *et al.*, 1999; Prince *et al.*, 2008; Saleh *et al.*, 2000; Shen *et al.*, 2004) and, in some cases, activation and repression domains have been mapped in Hox proteins (Rambaldi *et al.*, 1994; Tour, 2005; Viganò *et al.*, 1998; Zhao *et al.*, 1996). Covalent modifications, such as phosphorylation, have also been shown to influence Hox activities in interesting ways (Berry and Gehring, 2000; Galant and Carroll, 2002; Jaffe *et al.*, 1997; Ronshaugen *et al.*, 2002; Taghli-Lamallem *et al.*, 2008; reviewed elsewhere by Pearson *et al.*, 2005).

In addition, to these mechanisms, there are some recent examples suggesting that Hoxasomes may regulate transcription—and be regulated themselves—in other, mechanistically distinct ways. One example concerns

Author's personal copy

the way in which *Drosophila* Abd-A activates the expression of the gene *rho* in the peripheral nervous system (PNS). Abd-A-dependent expression of *rho*, which encodes a protease required for the processing of a ligand for the epidermal growth factor (EGF) receptor pathway, is necessary for abdominal-specific cell types (Brodu *et al.*, 2002). Abd-A carries out this paralog-specific function by assembling a multiprotein complex that includes both Exd and Hth cofactors (Gebelein and Mann, 2007). However, instead of the Abd-A Hoxasome directly activating *rho* transcription, it appears that it functions by competing with the binding of another transcription factor, Senseless, which is a *rho* repressor. Thus, the architecture of the *rho* *cis*-regulatory element is organized in a manner whereby the sequence-specific binding of an Abd-A Hoxasome permits *rho* expression by blocking binding of the Senseless repressor. Perhaps analogously, a zinc finger protein called ZFPIP has been shown to bind to Pbx1 and inhibit the binding of Hoxa9-Pbx1 complexes to a Hox-PBC consensus site (Laurent *et al.*, 2007). Thus, competition in DNA binding, rather than a direct influence on transcription, may underlie other examples of gene regulation by Hox factors.

A second interesting example of the importance of *cis*-regulatory element architecture comes from the analysis of an element from the *Drosophila bric-a-brac* (*bab*) gene, which is a direct target of Abd-B (Williams *et al.*, 2008). The activity of this element in the fifth and sixth abdominal segments of females—but not males—is critical for the dimorphic nature of abdominal pigmentation in male and female flies (Williams *et al.*, 2008). As Abd-B expression is the same in male and female flies, the sex-specific activities of this element stem from the Hox collaborator, *doublesex* (*dsx*), which is a downstream effector in the sex-determination pathway (Christiansen *et al.*, 2002). *dsx* encodes both male and female-specific isoforms. In males, the Dsx-M isoform collaborates with Abd-B to repress *bab*, while in females, the Dsx-F isoform collaborates with Abd-B to activate this *bab* element. In this element, there are two required Bab-binding sites, and more than 15 Abd-B-binding sites, suggesting the existence an unusually Hox-dense Hoxasome (Table 3.1). Moreover, the analysis of the same *cis*-regulatory element from other *Drosophila* species in which abdominal pigmentation pattern is the same in males and females suggests that the dimorphic activity of the *D. melanogaster bab* element is due to the relative orientation and specific spacing of the Dsx and Abd-B-binding sites. Mechanistically, it is currently unclear if these changes affect the stable assembly of this Hoxasome or, alternatively, its ability to recruit transcriptional coactivators.

Both of the studies highlighted above emphasize the value in characterizing bona fide Hox-targeted *cis*-regulatory elements at high resolution. Moreover, they also make it clear that a complete understanding of gene regulation by Hox proteins not only depends on understanding how these transcription factors bind DNA, but also how the bound factors, together

Author's personal copy

with their cofactors and collaborators, assemble and regulate transcription. We suspect that the discovery of additional regulatory mechanisms will depend on similar fine-scale analysis of other Hox-targeted *cis*-regulatory elements.

9. Conclusions

In this review, we have summarized a wide range of mechanisms that Hox proteins employ to regulate their target genes. For one, Hox proteins often require cofactors to bind to their binding sites in paralog-specific and semi-paralog-specific target genes. Cofactors may not be as essential, however, for shared Hox functions or those executed by Hox proteins in a unique regulatory environment, such as the *Drosophila* haltere. Structural studies have suggested that TALE family cofactors not only increase the size of the binding site, they help to impose additional structure onto otherwise unstructured homeodomain and nonhomeodomain residues, allowing them to read additional features present in Hox-cofactor-binding sites. It will be interesting to see how generally applicable this model for Hox-DNA binding (and perhaps other homeodomain proteins) will be as more PBC-Hox-DNA complexes are characterized at high resolution. Finally, we have also seen that Hox-regulated *cis*-regulatory elements utilize a potentially large number of protein collaborators, such as effector transcription factors that are downstream of cell-cell signaling pathways. The assembly of these multiprotein-DNA complexes, which we have called Hoxasomes to emphasize the central importance of the Hox input, is essential for dictating the sign (repression or activation) of the transcriptional regulation.

ACKNOWLEDGMENTS

We thank Matthew Slattery for comments on the manuscript and Barry Honig for discussions related to this review. This work was supported by an NIH R01 grant awarded to R.S.M.

REFERENCES

Affolter, M., Percival-Smith, A., Müller, M., Leupin, W., and Gehring, W. J. (1990). DNA binding properties of the purified Antennapedia homeodomain. *Proc. Natl. Acad. Sci. USA* **87,** 4093–4097.

Affolter, M., Slattery, M., and Mann, R. S. (2008). A lexicon for homeodomain-DNA recognition. *Cell* **133,** 1133–1135.

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell* **57,** 347–349.

Author's personal copy

Appel, B., and Sakonju, S. (1993). Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in Drosophila embryos. *EMBO J.* **12,** 1099–1109.

Arata, Y., Kouike, H., Zhang, Y., Herman, M. A., Okano, H., and Sawa, H. (2006). Wnt signaling and a Hox protein cooperatively regulate psa-3/Meis to determine daughter cell fate after asymmetric cell division in C. *elegans*. *Dev. Cell* **11,** 105–115.

Bandyopadhyay, S., Ashraf, M. Z., Daher, P., Howe, P. H., and DiCorleto, P. E. (2007). HOXA9 participates in the transcriptional activation of E-selectin in endothelial cells. *Mol. Cell Biol.* **27,** 4207–4216.

Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña-Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T., Khalid, F., Zhang, W., *et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133,** 1266–1276.

Berry, M., and Gehring, W. (2000). Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phosphatase 2A,B'. *EMBO J.* **19,** 2946–2957.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes. Dev.* **13,** 946–953.

Bessa, J., Tavares, M. J., Santos, J., Kikuta, H., Laplante, M., Becker, T. S., Gómez-Skarmeta, J. L., and Casares, F. (2008). meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. *Development* **135,** 799–803.

Bondos, S. E., and Tan, X. X. (2001). Combinatorial transcriptional regulation: the interaction of transcription factors and cell signaling molecules with homeodomain proteins in Drosophila development. *Crit. Rev. Eukaryot. Gene. Expr.* **11,** 145–171.

Bradley, P. L., Haberman, A. S., and Andrew, D. J. (2001). Organ formation in Drosophila: specification and morphogenesis of the salivary gland. *Bioessays* **23,** 901–911.

Brodu, V., Elstob, P. R., and Gould, A. P. (2002). abdominal A specifies one cell type in Drosophila by regulating one principal target gene. *Development* **129,** 2957–2963.

Bromleigh, V. C., and Freedman, L. P. (2000). p21 is a transcriptional target of HOXA10 in differentiating myelomonocytic cells. *Genes. Dev.* **14,** 2581–2586.

Bruhl, T., Urbich, C., Aicher, D., Acker-Palmer, A., Zeiher, A. M., and Dimmeler, S. (2004). Homeobox A9 transcriptionally regulates the EphB4 receptor to modulate endothelial cell migration and tube formation. *Circ. Res.* **94,** 743–751.

Capovilla, M., and Botas, J. (1998). Functional dominance among Hox genes: repression dominates activation in the regulation of Dpp. *Development* **125,** 4949–4957.

Capovilla, M., Brandt, M., and Botas, J. (1994). Direct regulation of decapentaplegic by Ultrabithorax and its role in Drosophila midgut morphogenesis. *Cell* **76,** 461–475.

Capovilla, M., Kambris, Z., and Botas, J. (2001). Direct regulation of the muscle-identity gene apterous by a Hox protein in the somatic mesoderm. *Development* **128,** 1221–1230.

Caré, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. (1996). HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell Biol.* **16,** 4842–4851.

Casares, F., Calleja, M., and Sánchez-Herrero, E. (1996). Functional similarity in appendage specification by the Ultrabithorax and abdominal-A Drosophila HOX genes. *EMBO J.* **15,** 3934–3942.

Casares, F., and Mann, R. S. (1998). Control of antennal versus leg development in Drosophila. *Nature* **392,** 723–726.

Casares, F., and Mann, R. S. (2001). The ground state of the ventral appendage in Drosophila. *Science* **293,** 1477–1480.

Author's personal copy

Ceccarelli, E., McGrew, M. J., Nguyen, T., Grieshammer, U., Horgan, D., Hughes, S. H., and Rosenthal, N. (1999). An E box comprises a positional sensor for regional differences in skeletal muscle gene expression and methylation. *Dev. Biol.* **213,** 217–229.

Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* **78,** 603–615.

Chan, S. K., and Mann, R. S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. *Genes Dev.* **7,** 796–811.

Chan, S. K., and Mann, R. S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. *Proc. Natl. Acad. Sci. USA* **93,** 5223–5228.

Chan, S. K., Ryoo, H. D., Gould, A., Krumlauf, R., and Mann, R. S. (1997). Switching the *in vivo* specificity of a minimal Hox-responsive element. *Development* **124,** 2007–2014.

Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C., and Cleary, M. L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **9,** 663–674.

Chariot, A., van Lint, C., Chapelier, M., Gielen, J., Merville, M. P., and Bours, V. (1999). CBP and histone deacetylase inhibition enhance the transactivation potential of the HOXB7 homeodomain-containing protein. *Oncogene* **18,** 4007–4014.

Chen, H., Rubin, E., Zhang, H., Chung, S., Jie, C. C., Garrett, E., Biswal, S., and Sukumar, S. (2005). Identification of transcriptional targets of HOXA5. *J. Biol. Chem.* **280,** 19373–19380.

Chen, J., and Ruley, H. E. (1998). An enhancer element in the EphA2 (Eck) gene sufficient for rhombomere-specific expression is activated by HOXA1 and HOXB1 homeobox proteins. *J. Biol. Chem.* **273,** 24670–24675.

Christiansen, A. E., Keisman, E. L., Ahmad, S. M., and Baker, B. S. (2002). Sex comes in from the cold: the integration of sex and pattern. *Trends Genet.* **18,** 510–516.

Coiffier, D., Charroux, B., and Kerridge, S. (2008). Common functions of central and posterior Hox genes for the repression of head in the trunk of Drosophila. *Development* **135,** 291–300.

Crickmore, M. A., and Mann, R. S. (2007). Hox control of morphogen mobility and organ development through regulation of glypican expression. *Development* **134,** 327–334.

Crickmore, M. A., and Mann, R. S. (2008). The control of size in animals: insights from selector genes. *Bioessays* **30,** 843–853.

Cui, M., and Han, M. (2003). Cis regulatory requirements for vulval cell-specific expression of the Caenorhabditis elegans fibroblast growth factor gene egl-17. *Dev. Biol.* **257,** 104–116.

Curtiss, J., Halder, G., and Mlodzik, M. (2002). Selector and signalling molecules cooperate in organ patterning. *Nat. Cell Biol.* **4,** E48–E51.

Daftary, G. S., Troy, P. J., Bagot, C. N., Young, S. L., and Taylor, H. S. (2002). Direct regulation of beta3-integrin subunit gene expression by HOXA10 in endometrial cells. *Mol. Endocrinol.* **16,** 571–579.

Dasen, J. S., De Camilli, A., Wang, B., Tucker, P. W., and Jessell, T. M. (2008). Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. *Cell* **134,** 304–316.

Desplan, C., Theis, J., and O'Farrell, P. H. (1988). The sequence specificity of homeodomain-DNA interaction. *Cell* **54,** 1081–1090.

Dessain, S., Gross, C. T., Kuziora, M. A., and McGinnis, W. (1992). Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J.* **11,** 991–1002.

Author's personal copy

Di Rocco, G., Gavalas, A., Popperl, H., Krumlauf, R., Mavilio, F., and Zappavigna, V. (2001). The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. *J. Biol. Chem.* **276,** 20506–20515.

Ebner, A., Cabernard, C., Affolter, M., and Merabet, S. (2005). Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. *Development* **132,** 1591–1600.

Edelman, G. M., and Jones, F. S. (1995). Developmental control of N-CAM expression by Hox and Pax gene products. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **349,** 305–312.

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *EMBO J.* **13,** 3551–3560.

Ekker, S. C., Young, K. E., von Kessler, D. P., and Beachy, P. A. (1991). Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. *EMBO J.* **10,** 1179–1186.

Estrada, B., and Sánchez-Herrero, E. (2001). The Hox gene Abdominal-B antagonizes appendage development in the genital disc of Drosophila. *Development* **128,** 331–339.

Ferretti, E., Cambronero, F., Tümpel, S., Longobardi, E., Wiedemann, L. M., Blasi, F., and Krumlauf, R. (2005). Hoxb1 enhancer and control of rhombomere 4 expression: complex interplay between PREP1-PBX1-HOXB1 binding sites. *Mol. Cell Biol.* **25,** 8541–8552.

Ferretti, E., Marshall, H., Pöpperl, H., Maconochie, M., Krumlauf, R., and Blasi, F. (2000). Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. *Development* **127,** 155–166.

Furukubo-Tokunaga, K., Flister, S., and Gehring, W. J. (1993). Functional specificity of the Antennapedia homeodomain. *Proc. Natl. Acad. Sci. USA* **90,** 6360–6364.

Galant, R., and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415,** 910–913.

Galant, R., Walsh, C. M., and Carroll, S. B. (2002). Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129,** 3115–3126.

García-Bellido, A. (1975). Genetic control of wing disc development in Drosophila. *Ciba Found Symp.* **0,** 161–182.

Gebelein, B., Culi, J., Ryoo, H. D., Zhang, W., and Mann, R. S. (2002). Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev. Cell* **3,** 487–498.

Gebelein, B., and Mann, R. (2007). Compartmental modulation of abdominal Hox expression by engrailed and sloppy-paired patterns the fly ectoderm. *Dev. Biol.* **308,** 593–605.

Gebelein, B., McKay, D. J., and Mann, R. S. (2004). Direct integration of Hox and segmentation gene inputs during Drosophila development. *Nature* **431,** 653–659.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell* **78,** 211–223.

Gibson, G., Schier, A., LeMotte, P., and Gehring, W. J. (1990). The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell* **62,** 1087–1103.

Glenn, S. T., Jones, C. A., Pan, L., and Gross, K. W. (2008). *In Vivo* Analysis of Key Elements within the Renin Regulatory Region. *Physiol. Genomics.* **35**(3), 243–253.

Gong, K. Q., Yallowitz, A. R., Sun, H., Dressler, G. R., and Wellik, D. M. (2007). A Hox-Eya-Pax complex regulates early kidney developmental gene expression. *Mol. Cell Biol.* **27,** 7661–7668.

Author's personal copy

Goodman, F. R. (2002). Limb malformations and the human HOX genes. *Am. J. Med. Genet.* **112,** 256–265.

Gould, A., Morrison, A., Sproat, G., White, R. A., and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev.* **11,** 900–913.

Greig, S., and Akam, M. (1995). The role of homeotic genes in the specification of the Drosophila gonad. *Curr. Biol.* **5,** 1057–1062.

Grienenberger, A., Merabet, S., Manak, J., Iltis, I., Fabre, A., Bérenger, H., Scott, M. P., Pradel, J., and Graba, Y. (2003). Tgfbeta signaling acts on a Hox response element to confer specificity and diversity to Hox protein function. *Development* **130,** 5445–5455.

Guazzi, S., Lonigro, R., Pintonello, L., Boncinelli, E., Di Lauro, R., and Mavilio, F. (1994). The thyroid transcription factor-1 gene is a candidate target for regulation by Hox proteins. *EMBO J.* **13,** 3339–3347.

Haerry, T. E., and Gehring, W. J. (1997). A conserved cluster of homeodomain binding sites in the mouse Hoxa-4 intron functions in Drosophila embryos as an enhancer that is directly regulated by Ultrabithorax. *Dev. Biol.* **186,** 1–15.

Haller, K., Rambaldi, I., Daniels, E., and Featherstone, M. (2004). Subcellular localization of multiple PREP2 isoforms is regulated by actin, tubulin, and nuclear export. *J. Biol. Chem.* **279,** 49384–49394.

Hassan, M. Q., Tare, R., Lee, S. H., Mandeville, M., Weiner, B., Montecino, M., van Wijnen, A. J., Stein, J. L., Stein, G. S., and Lian, J. B. (2007). HOXA10 controls osteoblastogenesis by directly activating bone regulatory and phenotypic genes. *Mol. Cell Biol.* **27,** 3337–3352.

Hersh, B. M., and Carroll, S. B. (2005). Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in Drosophila. *Development* **132,** 1567–1577.

Hersh, B. M., Nelson, C. E., Stoll, S. J., Norton, J. E., Albert, T. J., and Carroll, S. B. (2007). The UBX-regulated network in the haltere imaginal disc of D. melanogaster. *Dev. Biol.* **302,** 717–727.

Hirth, F., Loop, T., Egger, B., Miller, D. F., Kaufman, T. C., and Reichert, H. (2001). Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of Drosophila. *Development* **128,** 4781–4788.

Hoey, T., and Levine, M. (1988). Divergent homeo box proteins recognize similar DNA sequences in Drosophila. *Nature* **332,** 858–861.

Houghton, L., and Rosenthal, N. (1999). Regulation of a muscle-specific transgene by persistent expression of Hox genes in postnatal murine limb muscle. *Dev. Dyn.* **216,** 385–397.

Huang, H., Paliouras, M., Rambaldi, I., Lasko, P., and Featherstone, M. (2003). Nonmuscle myosin promotes cytoplasmic localization of PBX. *Mol. Cell Biol.* **23,** 3636–3645.

Hueber, S. D., Bezdan, D., Henz, S. R., Blank, M., Wu, H., and Lohmann, I. (2007). Comparative analysis of Hox downstream genes in Drosophila. *Development* **134,** 381–392.

Hueber, S. D., and Lohmann, I. (2008). Shaping segments: Hox gene function in the genomic age. *Bioessays* **30,** 965–979.

Jaffe, L., Ryoo, H. D., and Mann, R. S. (1997). A role for phosphorylation by casein kinase II in modulating Antennapedia activity in Drosophila. *Genes Dev.* **11,** 1327–1340.

Jave-Suárez, L. F., and Schweizer, J. (2006). The HOXC13-controlled expression of early hair keratin genes in the human hair follicle does not involve TALE proteins MEIS and PREP as cofactors. *Arch. Dermatol. Res.* **297,** 372–376.

Jave-Suarez, L. F., Winter, H., Langbein, L., Rogers, M. A., and Schweizer, J. (2002). HOXC13 is involved in the regulation of human hair keratin gene expression. *J. Biol. Chem.* **277,** 3718–3726.

Author's personal copy

Jeong, S., Rokas, A., and Carroll, S. B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell* **125,** 1387–1399.

Jiang, Y., Shi, H., Amin, N. M., Sultan, I., and Liu, J. (2008). Mesodermal expression of the *C. elegans* HMX homolog mls-2 requires the PBC homolog CEH-20. *Mech. Dev.* **125,** 451–461.

Jiménez, G., Paroush, Z., and Ish-Horowicz, D. (1997). Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev.* **11,** 3072–3082.

Jones, F. S., Holst, B. D., Minowa, O., De Robertis, E. M., and Edelman, G. M. (1993). Binding and transcriptional activation of the promoter for the neural cell adhesion molecule by HoxC6 (Hox-3.3). *Proc. Natl. Acad. Sci. USA* **90,** 6557–6561.

Jones, F. S., Prediger, E. A., Bittner, D. A., De Robertis, E. M., and Edelman, G. M. (1992). Cell adhesion molecules as targets for Hox genes: neural cell adhesion molecule promoter activity is modulated by cotransfection with Hox-2.5 and -2.4. *Proc. Natl. Acad. Sci. USA* **89,** 2086–2090.

Joshi, R., Passner, J., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M., Jacob, V., Aggarwal, A., Honig, B., and Mann, R. (2007). Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. *Cell* **131,** 530–543.

Joulia, L., Deutsch, J., Bourbon, H. M., and Cribbs, D. L. (2006). The specification of a highly derived arthropod appendage, the Drosophila labial palps, requires the joint action of selectors and signaling pathways. *Dev. Genes Evol.* **216,** 431–442.

Kataoka, K., Yoshitomo-Nakagawa, K., Shioda, S., and Nishizawa, M. (2001). A set of Hox proteins interact with the Maf oncoprotein to inhibit its DNA binding, transactivation, and transforming activities. *J. Biol. Chem.* **276,** 819–826.

Knosp, W. M., Saneyoshi, C., Shou, S., Bächinger, H. P., and Stadler, H. S. (2007). Elucidation, quantitative refinement, and *in vivo* utilization of the HOXA13 DNA binding site. *J. Biol. Chem.* **282,** 6843–6853.

Koh, K., Peyrot, S. M., Wood, C. G., Wagmaister, J. A., Maduro, M. F., Eisenmann, D. M., and Rothman, J. H. (2002). Cell fates and fusion in the *C. elegans* vulval primordium are regulated by the EGL-18 and ELT-6 GATA factors -- apparent direct targets of the LIN-39 Hox protein. *Development* **129,** 5171–5180.

Kornfeld, K., Saint, R. B., Beachy, P. A., Harte, P. J., Peattie, D. A., and Hogness, D. S. (1989). Structure and expression of a family of Ultrabithorax mRNAs generated by alternative splicing and polyadenylation in Drosophila. *Genes Dev.* **3,** 243–258.

Kremser, T., Hasenpusch-Theil, K., Wagner, E., Buttgereit, D., and Renkawitz-Pohl, R. (1999). Expression of the beta3 tubulin gene (beta Tub60D) in the visceral mesoderm of Drosophila is dependent on a complex enhancer that binds Tinman and UBX. *Mol. Gen. Genet.* **262,** 643–658.

Kuziora, M. A., and McGinnis, W. (1989). A homeodomain substitution changes the regulatory specificity of the deformed protein in Drosophila embryos. *Cell* **59,** 563–571.

Kuziora, M. A., and McGinnis, W. (1990). Altering the regulatory targets of the Deformed protein in Drosophila embryos by substituting the Abdominal-B homeodomain. *Mech. Dev.* **33,** 83–93.

Lampe, X., Picard, J. J., and Rezsohazy, R. (2004). The Hoxa2 enhancer 2 contains a critical Hoxa2 responsive regulatory element. *Biochem. Biophys. Res. Commun.* **316,** 898–902.

Lampe, X., Samad, O. A., Guiguen, A., Matis, C., Remacle, S., Picard, J. J., Rijli, F. M., and Rezsohazy, R. (2008). An ultraconserved Hox-Pbx responsive element resides in the coding sequence of Hoxa2 and is active in rhombomere 4. *Nucleic Acids Res.* **36,** 3214–3225.

LaRonde-LeBlanc, N. A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev.* **17,** 2060–2072.

Author's personal copy

Laurent, A., Bihan, R., Deschamps, S., Guerrier, D., Dupé, V., Omilli, F., Burel, A., and Pellerin, I. (2007). Identification of a new type of PBX1 partner that contains zinc finger motifs and inhibits the binding of HOXA9-PBX1 to DNA. *Mech. Dev.* **124,** 364–376.

Lei, H., Wang, H., Juan, A. H., and Ruddle, F. H. (2005). The identification of Hoxc8 target genes. *Proc. Natl. Acad. Sci. USA* **102,** 2420–2424.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276,** 565–570.

Li-Kroeger, D., Witt, L. M., Grimes, H. L., Cook, T. A., and Gebelein, B. (2008). Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the Drosophila PNS. *Dev. Cell* **15,** 298–308.

Lin, L., and McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. *Genes Dev.* **6,** 1071–1081.

Liu, H., Strauss, T. J., Potts, M. B., and Cameron, S. (2006). Direct regulation of egl-1 and of programmed cell death by the Hox protein MAB-5 and by CEH-20, a *C. elegans* homolog of Pbx1. *Development* **133,** 641–650.

Liu, J., and Fire, A. (2000). Overlapping roles of two Hox genes and the exd ortholog ceh-20 in diversification of the *C. elegans* postembryonic mesoderm. *Development* **127,** 5179–5190.

Lohmann, I., McGinnis, N., Bodmer, M., and McGinnis, W. (2002). The Drosophila Hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. *Cell* **110,** 457–466.

Lu, Q., and Kamps, M. P. (1996). Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a model of a Pbx1-Hox-DNA complex. *Mol. Cell Biol.* **16,** 1632–1640.

Luo, L., Yang, X., Takihara, Y., Knoetgen, H., and Kessel, M. (2004). The cell-cycle regulator geminin inhibits Hox function through direct and polycomb-mediated interactions. *Nature* **427,** 749–753.

Manak, J. R., Mathies, L. D., and Scott, M. P. (1994). Regulation of a decapentaplegic midgut enhancer by homeotic proteins. *Development* **120,** 3605–3619.

Mann, R. S. (1995). The specificity of homeotic gene function. *Bioessays* **17,** 855–863.

Mann, R. S., and Abu-Shaar, M. (1996). Nuclear import of the homeodomain protein extradenticle in response to Wg and Dpp signalling. *Nature* **383,** 630–633.

Mann, R. S., and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8,** 423–429.

Mann, R. S., and Carroll, S. B. (2002). Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* **12,** 592–600.

Mann, R. S., and Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet.* **12,** 258–262.

Mann, R. S., and Hogness, D. S. (1990). Functional dissection of Ultrabithorax proteins in D. melanogaster. *Cell* **60,** 597–610.

Mann, R. S., and Morata, G. (2000). The developmental and molecular biology of genes that subdivide the body of Drosophila. *Annu. Rev. Cell Dev. Biol.* **16,** 243–271.

Manzanares, M., Bel-Vialar, S., Ariza-McNaughton, L., Ferretti, E., Marshall, H., Maconochie, M. M., Blasi, F., and Krumlauf, R. (2001). Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms. *Development* **128,** 3595–3607.

Marty, T., Vigano, M. A., Ribeiro, C., Nussbaumer, U., Grieder, N. C., and Affolter, M. (2001). A HOX complex, a repressor element and a 50 bp sequence confer regional specificity to a DPP-responsive enhancer. *Development* **128,** 2833–2845.

McCabe, C. D., and Innis, J. W. (2005). A genomic approach to the identification and characterization of HOXA13 functional binding elements. *Nucleic Acids Res.* **33,** 6782–6794.

Author's personal copy

McCormick, A., Coré, N., Kerridge, S., and Scott, M. P. (1995). Homeotic response elements are tightly linked to tissue-specific elements in a transcriptional enhancer of the teashirt gene. *Development* **121,** 2799–2812.

Merabet, S., Hombria, J. C., Hu, N., Pradel, J., and Graba, Y. (2005). Hox-controlled reorganisation of intrasegmental patterning cues underlies Drosophila posterior spiracle organogenesis. *Development* **132,** 3093–3102.

Merabet, S., Kambris, Z., Capovilla, M., Bérenger, H., Pradel, J., and Graba, Y. (2003). The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. *Dev. Cell* **4,** 761–768.

Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., and Graba, Y. (2007). A unique Extradenticle recruitment mode in the Drosophila Hox protein Ultrabithorax. *Proc. Natl. Acad. Sci. USA* **104,** 16946–16951.

Milech, N., Kees, U. R., and Watt, P. M. (2001). Novel alternative PBX3 isoforms in leukemia cells with distinct interaction specificities. *Genes Chromosomes Cancer* **32,** 275–780.

Moens, C. B., and Selleri, L. (2006). Hox cofactors in vertebrate development. *Dev. Biol.* **291,** 193–206.

Monica, K., Galili, N., Nourse, J., Saltman, D., and Cleary, M. L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Mol. Cell Biol.* **11,** 6149–6157.

Morgan, R., In der Rieden, P., Hooiveld, M. H., and Durston, A. J. (2000). Identifying HOX paralog groups by the PBX-binding region. *Trends Genet.* **16,** 66–67.

Mukherjee, K., and Bürglin, T. R. (2007). Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. *J. Mol. Evol.* **65,** 137–153.

Nakano, T., Windrem, M., Zappavigna, V., and Goldman, S. A. (2005). Identification of a conserved 125 base-pair Hb9 enhancer that specifies gene expression to spinal motor neurons. *Dev. Biol.* **283,** 474–485.

Neuteboom, S. T., Peltenburg, L. T., van Dijk, M. A., and Murre, C. (1995). The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins. *Proc. Natl. Acad. Sci. USA* **92,** 9166–9170.

Noro, B., Culi, J., McKay, D. J., Zhang, W., and Mann, R. S. (2006). Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by homothorax. *Genes Dev.* **20,** 1636–1650.

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., and Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133,** 1277–1289.

O'Connor, M. B., Binari, R., Perkins, L. A., and Bender, W. (1988). Alternative RNA products from the Ultrabithorax domain of the bithorax complex. *EMBO J.* **7,** 435–445.

Pan, L., Glenn, S. T., Jones, C. A., and Gross, K. W. (2005). Activation of the rat renin promoter by HOXD10.PBX1b.PREP1, Ets-1, and the intracellular domain of notch. *J. Biol. Chem.* **280,** 20860–20866.

Pan, L., Xie, Y., Black, T. A., Jones, C. A., Pruitt, S. C., and Gross, K. W. (2001). An Abd-B class HOX.PBX recognition sequence is required for expression from the mouse Ren-1c gene. *J. Biol. Chem.* **276,** 32489–32494.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* **397,** 714–719.

Pearson, J. C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6,** 893–904.

Pederson, J. A., LaFollette, J. W., Gross, C., Veraksa, A., McGinnis, W., and Mahaffey, J. W. (2000). Regulation by homeoproteins: a comparison of deformed-responsive elements. *Genetics* **156,** 677–686.

Author's personal copy

Penkov, D., Tanaka, S., Di Rocco, G., Berthelsen, J., Blasi, F., and Ramirez, F. (2000). Cooperative interactions between PBX, PREP, and HOX proteins modulate the activity of the alpha 2(V) collagen (COL5A2) promoter. *J. Biol. Chem.* **275,** 16681–16689.

Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell Biol.* **15,** 3989–3997.

Pinsonneault, J., Florence, B., Vaessin, H., and McGinnis, W. (1997). A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. *EMBO J.* **16,** 2032–2042.

Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96,** 587–597.

Plaza, S., Prince, F., Adachi, Y., Punzo, C., Cribbs, D. L., and Gehring, W. J. (2008). Cross-regulatory protein-protein interactions between Hox and Pax transcription factors. *Proc. Natl. Acad. Sci. USA* **105,** 13439–13444.

Pöpperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S., and Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell* **81,** 1031–1042.

Potter, C. S., Peterson, R. L., Barth, J. L., Pruett, N. D., Jacobs, D. F., Kern, M. J., Argraves, W. S., Sundberg, J. P., and Awgulewitsch, A. (2006). Evidence that the satin hair mutant gene Foxq1 is among multiple and functionally diverse regulatory targets for Hoxc13 during hair follicle differentiation. *J. Biol. Chem.* **281,** 29245–29255.

Prévôt, D., Voeltzel, T., Birot, A. M., Morel, A. P., Rostan, M. C., Magaud, J. P., and Corbo, L. (2000). The leukemia-associated protein Btg1 and the p53-regulated protein Btg2 interact with the homeoprotein Hoxb9 and enhance its transcriptional activation. *J. Biol. Chem.* **275,** 147–153.

Prince, F., Katsuyama, T., Oshima, Y., Plaza, S., Resendez-Perez, D., Berry, M., Kurata, S., and Gehring, W. J. (2008). The YPWM motif links Antennapedia to the basal transcriptional machinery. *Development* **135,** 1669–1679.

Raman, V., Martensen, S. A., Reisman, D., Evron, E., Odenwald, W. F., Jaffee, E., Marks, J., and Sukumar, S. (2000a). Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405,** 974–978.

Raman, V., Tamori, A., Vali, M., Zeller, K., Korz, D., and Sukumar, S. (2000b). HOXA5 regulates expression of the progesterone receptor. *J. Biol. Chem.* **275,** 26551–26555.

Rambaldi, I., Kovàcs, E. N., and Featherstone, M. S. (1994). A proline-rich transcriptional activation domain in murine HOXD-4 (HOX-4.2). *Nucleic Acids Res.* **22,** 376–382.

Randazzo, F. M., Cribbs, D. L., and Kaufman, T. C. (1991). Rescue and regulation of proboscipedia: a homeotic gene of the Antennapedia Complex. *Development* **113,** 257–271.

Rao, M. V., Donoghue, M. J., Merlie, J. P., and Sanes, J. R. (1996). Distinct regulatory elements control muscle-specific, fiber-type-selective, and axially graded expression of a myosin light-chain gene in transgenic mice. *Mol. Cell Biol.* **16,** 3909–3922.

Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M., and McGinnis, W. (1985). Homeo box genes of the Antennapedia and bithorax complexes of Drosophila. *Cell* **43,** 71–80.

Robertson, L. K., Bowling, D. B., Mahaffey, J. P., Imiolczyk, B., and Mahaffey, J. W. (2004). An interactive network of zinc-finger proteins contributes to regionalization of the Drosophila embryo and establishes the domains of HOM-C protein function. *Development* **131,** 2781–2789.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* **415,** 914–917.

Rousso, D. L., Gaber, Z. B., Wellik, D., Morrisey, E. E., and Novitch, B. G. (2008). Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. *Neuron* **59,** 226–240.

Author's personal copy

Ryoo, H. D., and Mann, R. S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* **13,** 1704–1716.

Ryoo, H. D., Marty, T., Casares, F., Affolter, M., and Mann, R. S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. *Development* **126,** 5137–5148.

Saleh, M., Rambaldi, I., Yang, X. J., and Featherstone, M. S. (2000). Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. *Mol. Cell Biol.* **20,** 8623–8633.

Salsi, V., and Zappavigna, V. (2006). Hoxd13 and Hoxa13 directly control the expression of the EphA7 Ephrin tyrosine kinase receptor in developing limbs. *J. Biol. Chem.* **281,** 1992–1999.

Samad, O. A., Geisen, M. J., Caronia, G., Varlet, I., Zappavigna, V., Ericson, J., Goridis, C., and Rijli, F. M. (2004). Integration of anteroposterior and dorsoventral regulation of Phox2b transcription in cranial motoneuron progenitors by homeodomain proteins. *Development* **131,** 4071–4083.

Serpente, P., Tümpel, S., Ghyselinck, N. B., Niederreither, K., Wiedemann, L. M., Dollé, P., Chambon, P., Krumlauf, R., and Gould, A. P. (2005). Direct crossregulation between retinoic acid receptor {beta} and Hox genes during hindbrain segmentation. *Development* **132,** 503–513.

Shen, W., Chrobak, D., Krishnan, K., Lawrence, H. J., and Largman, C. (2004). HOXB6 protein is bound to CREB-binding protein and represses globin expression in a DNA binding-dependent, PBX interaction-independent process. *J. Biol. Chem.* **279,** 39895–39904.

Shen, W. F., Rozenfeld, S., Lawrence, H. J., and Largman, C. (1997). The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. *J. Biol. Chem.* **272,** 8198–81206.

Shi, X., Bai, S., Li, L., and Cao, X. (2001). Hoxa-9 represses transforming growth factor-beta-induced osteopontin gene transcription. *J. Biol. Chem.* **276,** 850–855.

Shi, X., Yang, X., Chen, D., Chang, Z., and Cao, X. (1999). Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J. Biol. Chem.* **274,** 13711–13717.

Stevens, K. E., and Mann, R. S. (2007). A balance between two nuclear localization sequences and a nuclear export sequence governs extradenticle subcellular localization. *Genetics* **175,** 1625–1636.

Streit, A., Kohler, R., Marty, T., Belfiore, M., Takacs-Vellai, K., Vigano, M. A., Schnabel, R., Affolter, M., and Müller, F. (2002). Conserved regulation of the Caenorhabditis elegans labial/Hox1 gene ceh-13. *Dev. Biol.* **242,** 96–108.

Stultz, B. G., Jackson, D. G., Mortin, M. A., Yang, X., Beachy, P. A., and Hursh, D. A. (2006). Transcriptional activation by extradenticle in the Drosophila visceral mesoderm. *Dev. Biol.* **290,** 482–494.

Sun, B., Hursh, D. A., Jackson, D., and Beachy, P. A. (1995). Ultrabithorax protein is necessary but not sufficient for full activation of decapentaplegic expression in the visceral mesoderm. *EMBO J.* **14,** 520–535.

Taghli-Lamallem, O., Gallet, A., Leroy, F., Malapert, P., Vola, C., Kerridge, S., and Fasano, L. (2007). Direct interaction between Teashirt and Sex combs reduced proteins, via Tsh's acidic domain, is essential for specifying the identity of the prothorax in Drosophila. *Dev. Biol.* **307,** 142–151.

Taghli-Lamallem, O., Hsia, C., Ronshaugen, M., and McGinnis, W. (2008). Context-dependent regulation of Hox protein functions by CK2 phosphorylation sites. *Dev. Genes Evol.* **218,** 321–332.

Tour, E. (2005). Evolutionarily conserved domains required for activation and repression functions of the Drosophila Hox protein Ultrabithorax. *Development* **132,** 5271–5281.

Author's personal copy

Troy, P. J., Daftary, G. S., Bagot, C. N., and Taylor, H. S. (2003). Transcriptional repression of peri-implantation EMX2 expression in mammalian reproduction by HOXA10. *Mol. Cell Biol.* **23,** 1–13.

Tümpel, S., Cambronero, F., Ferretti, E., Blasi, F., Wiedemann, L. M., and Krumlauf, R. (2007). Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. *Dev. Biol.* **302,** 646–660.

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J., and Cohen, S. M. (1992). Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less. *Cell* **71,** 437–450.

Viganò, M. A., Di Rocco, G., Zappavigna, V., and Mavilio, F. (1998). Definition of the transcriptional activation domains of three human HOX proteins depends on the DNA-binding context. *Mol. Cell Biol.* **18,** 6201–6212.

Vincent, J. P., Kassis, J. A., and O'Farrell, P. H. (1990). A synthetic homeodomain binding site acts as a cell type specific, promoter specific enhancer in Drosophila embryos. *EMBO J.* **9,** 2573–2578.

Wagmaister, J. A., Miley, G. R., Morris, C. A., Gleason, J. E., Miller, L. M., Kornfeld, K., and Eisenmann, D. M. (2006). Identification of cis-regulatory elements from the *C. elegans* Hox gene lin-39 required for embryonic expression and for regulation by the transcription factors LIN-1, LIN-31 and LIN-39. *Dev. Biol.* **297,** 550–565.

Wagner, K., Mincheva, A., Korn, B., Lichter, P., and Popperl, H. (2001). Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mech. Dev.* **103,** 127–131.

Walsh, C. M., and Carroll, S. B. (2007). Collaboration between Smads and a Hox protein in target gene repression. *Development* **134,** 3585–3592.

Wang, N., Kim, H. G., Cotta, C. V., Wan, M., Tang, Y., Klug, C. A., and Cao, X. (2006). TGFbeta/BMP inhibits the bone marrow transformation capability of Hoxa9 by repressing its DNA-binding ability. *EMBO J.* **25,** 1469–1480.

Weatherbee, S. D., and Carroll, S. B. (1999). Selector genes and limb identity in arthropods and vertebrates. *Cell* **97,** 283–286.

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., and Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell* **134,** 610–623.

Williams, T. M., Williams, M. E., Heaton, J. H., Gelehrter, T. D., and Innis, J. W. (2005). Group 13 HOX proteins interact with the MH2 domain of R-Smads and modulate Smad transcriptional activation functions independent of HOX DNA-binding capability. *Nucleic Acids Res.* **33,** 4475–4484.

Wu, Y., Moser, M., Bautch, V. L., and Patterson, C. (2003). HoxB5 is an upstream transcriptional switch for differentiation of the vascular endothelium from precursor cells. *Mol. Cell Biol.* **23,** 5680–5691.

Yao, L. C., Liaw, G. J., Pai, C. Y., and Sun, Y. H. (1999). A common mechanism for antenna-to-Leg transformation in Drosophila: suppression of homothorax transcription by four HOM-C genes. *Dev. Biol.* **211,** 268–276.

Zeng, C., Pinsonneault, J., Gellon, G., McGinnis, N., and McGinnis, W. (1994). Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. *EMBO J.* **13,** 2362–2377.

Zhao, J. J., Lazzarini, R. A., and Pick, L. (1996). Functional dissection of the mouse Hox-a5 gene. *EMBO J.* **15,** 1313–1322.

Zhao, Y., and Potter, S. S. (2001). Functional specificity of the Hoxa13 homeobox. *Development* **128,** 3197–3207.

Zhao, Y., and Potter, S. S. (2002). Functional comparison of the Hoxa 4, Hoxa 10, and Hoxa 11 homeoboxes. *Dev. Biol.* **244,** 21–36.

Zhou, B., Chen, L., Wu, X., Wang, J., Yin, Y., and Zhu, G. (2008). MH1 domain of SMAD4 binds N-terminal residues of the homeodomain of Hoxc9. *Biochim. Biophys. Acta* **1784,** 747–752.